IJCSIS Vol. 16 No. 8, August 2018
ISSN 1947-5500

International Journal of Computer Science & Information Security

Pennsylvania, USA

Indexed and technically co-sponsored by :

Indexing Service

IJCSIS has been indexed by several world class databases, for more information, please access the following links:

Global Impact Factor
http://globalimpactfactor.com/
Google Scholar
http://scholar.google.com/
CrossRef
http://www.crossref.org/
Microsoft Academic Search
http://academic.research.microsoft.com/
IndexCopernicus
http://journals.indexcopernicus.com/
IET Inspec
http://www.theiet.org/resources/inspec/
EBSCO
http://www.ebscohost.com/
JournalSeek
http://journalseek.net
Ulrich
http://ulrichsweb.serialssolutions.com/
WordCat
http://www.worldcat.org
Academic Journals Database
http://www.journaldatabase.org/
Stanford University Libraries
http://searchworks.stanford.edu/
Harvard Library
http://discovery.lib.harvard.edu/?itemid=|library/m/aleph|012618581
UniSA Library
http://www.library.unisa.edu.au/
ProQuest
http://www.proquest.co.uk

Zeitschriftendatenbank (ZDB)
http://dispatch.opac.d-nb.de/

IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS
International Journal of Computer Science and Information Security (IJCSIS)
January-December 2018 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.
See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: see web site
Notification: see web site
Revision: see web site
Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies, IoT
Digital information processing
Cognitive science and knowledge
Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site https://sites.google.com/site/ijcsis/

For more information, please visit the journal website (https://sites.google.com/site/ijcsis/)

Editorial
Message from Editorial Board

*It is our great pleasure to present the **August 2018 issue** (Volume **16** Number **8**) of the **International Journal of Computer Science and Information Security** (IJCSIS). High quality research, survey & review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and digital technologies. It especially provides a platform for high-caliber academics, practitioners and PhD/Doctoral graduates to publish completed work and latest research outcomes. According to Google Scholar, up to now papers published in IJCSIS have been cited over 11450 times and this journal is experiencing steady and healthy growth. Google statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. There have been many improvements to the processing of papers; we have also witnessed a significant growth in interest through a higher number of submissions as well as through the breadth and quality of those submissions. IJCSIS is already indexed in some major academic/scientific databases and important repositories, such as: Google Scholar, Thomson Reuters, ArXiv, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, ResearchGate, LinkedIn, Academia.edu and EBSCO among others.*

A reputed & professional journal has a dedicated editorial team of editors and reviewers. On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers & editors for their outstanding efforts to meticulously review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing or reading papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status, making sure we deliver high-quality content to our readers in a timely fashion.

"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." We would like to thank you, the authors and readers, the content providers and consumers, who have made this journal the best possible.

*For further questions or other suggestions please do not hesitate to contact us at **ijcsiseditor@gmail.com**.*

A complete list of journals can be found at:
http://sites.google.com/site/ijcsis/

IJCSIS Vol. 16, No. 8, August 2018 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

Bibliographic Information
ISSN: 1947-5500
Monthly publication (Regular Special Issues)
Commenced Publication since May 2009

Editorial / Paper Submissions:
IJCSIS Managing Editor
(ijcsiseditor@gmail.com)
Pennsylvania, USA
Tel: +1 412 390 5159

IJCSIS EDITORIAL BOARD

IJCSIS Editorial Board	IJCSIS Guest Editors / Associate Editors
Dr. Shimon K. Modi [Profile] Director of Research BSPA Labs, Purdue University, USA	**Dr Riktesh Srivastava [Profile]** Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE
Professor Ying Yang, PhD. [Profile] Computer Science Department, Yale University, USA	**Dr. Jianguo Ding [Profile]** Norwegian University of Science and Technology (NTNU), Norway
Professor Hamid Reza Naji, PhD. [Profile] Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran	**Dr. Naseer Alquraishi [Profile]** University of Wasit, Iraq
Professor Yong Li, PhD. [Profile] School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China	**Dr. Kai Cong [Profile]** Intel Corporation, & Computer Science Department, Portland State University, USA
Professor Mokhtar Beldjehem, PhD. [Profile] Sainte-Anne University, Halifax, NS, Canada	**Dr. Omar A. Alzubi [Profile]** Al-Balqa Applied University (BAU), Jordan
Professor Yousef Farhaoui, PhD. Department of Computer Science, Moulay Ismail University, Morocco	**Dr. Jorge A. Ruiz-Vanoye [Profile]** Universidad Autónoma del Estado de Morelos, Mexico
Dr. Alex Pappachen James [Profile] Queensland Micro-nanotechnology center, Griffith University, Australia	**Prof. Ning Xu,** Wuhan University of Technology, China
Professor Sanjay Jasola [Profile] Gautam Buddha University	**Dr . Bilal Alatas [Profile]** Department of Software Engineering, Firat University, Turkey
Dr. Siddhivinayak Kulkarni [Profile] University of Ballarat, Ballarat, Victoria, Australia	**Dr. Ioannis V. Koskosas,** University of Western Macedonia, Greece
Dr. Reza Ebrahimi Atani [Profile] University of Guilan, Iran	**Dr Venu Kuthadi [Profile]** University of Johannesburg, Johannesburg, RSA
Dr. Dong Zhang [Profile] University of Central Florida, USA	**Dr. Zhihan lv [Profile]** Chinese Academy of Science, China
Dr. Vahid Esmaeelzadeh [Profile] Iran University of Science and Technology	**Prof. Ghulam Qasim [Profile]** University of Engineering and Technology, Peshawar, Pakistan
Dr. Jiliang Zhang [Profile] Northeastern University, China	**Prof. Dr. Maqbool Uddin Shaikh [Profile]** Preston University, Islamabad, Pakistan
Dr. Jacek M. Czerniak [Profile] Casimir the Great University in Bydgoszcz, Poland	**Dr. Musa Peker [Profile]** Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Binh P. Nguyen [Profile] National University of Singapore	**Dr. Wencan Luo [Profile]** University of Pittsburgh, US
Professor Seifeidne Kadry [Profile] American University of the Middle East, Kuwait	**Dr. Ijaz Ali Shoukat [Profile]** King Saud University, Saudi Arabia
Dr. Riccardo Colella [Profile] University of Salento, Italy	**Dr. Yilun Shang [Profile]** Tongji University, Shanghai, China
Dr. Sedat Akleylek [Profile] Ondokuz Mayis University, Turkey	**Dr. Sachin Kumar [Profile]** Indian Institute of Technology (IIT) Roorkee

ISSN 1947 5500

TABLE OF CONTENTS

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

17. PaperID 30041870: Information Extraction from Dependency Graphs (pp. 162-173)

A. Murali Krishna, Research scholar, Dept. of Computer Science & Engineering, Rayalaseema University Kurnool – 518007 . (A.P), INDIA.
Dr. T. Swarna Latha, Professor, Dept. of Computer Science & Engineering, Narayana Engineering College Nellore .AP, INDIA.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

18. PaperID 31071806: Object Detection in Fog Degraded Images (pp. 174-182)

Gurveer Singh & Dr. Ashima Singh
Thapar Institute of Engineering and Technology, Patiala, India

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

Online Security and Privacy Concerns: Issues and Recommendations

Binh Tran
School of Science and Technology
Georgia Gwinnett College
Lawrenceville, GA 30043, USA
Tel: 770.778.8542
Email: btran5@ggc.edu

Mirza B. Murtaza
School of Science and Technology
Georgia Gwinnett College
Lawrenceville, GA 30043, USA
Tel: 470.955.0589
Email: mmurtaza@ggc.edu

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Online Security and Privacy Concerns: Issues and Recommendations

Abstract

Security and privacy of information has been issues of significant concerns in last few decades both in personal lives and in the business world. Instances of security and privacy issues are widespread from individual's identity theft to major data breaches involving hundreds of thousands of customers at a time. The authors in this paper performed a survey analysis of more than 125 individuals and also conducted a phishing experiment to see how many of the students react to such emails. The conclusions include use of frequent employee training and maintaining processes in place that can ensure each member of the organization is aware of current security and privacy issues.

Keywords: information security, privacy, security issues, security breaches

Introduction

Information security or cyber security is increasingly becoming more important in both the business and the private world today especially with the popularity of mobile devices that connect the roughly 2.5 billion users worldwide. Nearly everyone agrees that cyber security is important and can pose serious threats; however, not everything is being done to help mitigate these threats from education, training and awareness. So often there are new reports of cyber security breaches at various companies such as Target, Home Depot, Equifax and most recently privacy concerns of Facebook. Corporate training has typically included short videos and acknowledgement of viewing them without proper assessment on whether these materials have clearly been understood to the detriment of the organization. Phishing, a form of electronic identity theft in which a combination of social engineering and website spoofing techniques are used to gain information that a user reveals is becoming so abundant. Training and education must keep up to date with the increasingly complex variations of cyber security. In this paper, the authors have created a survey to be completed by entry level students on privacy and security concerns as well as a phishing email experiment with results, conclusions and suggestions.

Privacy and Security Concerns

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The theft of personally identifiable information and data have been on the rise [1] and also there has been increase in financial fraud [2]. Security and privacy concerns can also be an hurdle in IoT (Internet of Things) adoption [3] Some research has been done on employee behaviors in the corporate environments, but there is a need to better understand of the online behavior of individual users, and to build awareness to avoid technology-enabled loss of privacy. The theoretical frameworks used to help explain online behavior include the theory of reasoned action and planned behavior to examine individuals' decisions when performing certain activities [4].

Privacy and security of consumers on the Internet has been under threat for a long time, several studies have indicated that users are concerned about the way privacy is handled in the information age. These concerns are not new, from the advent of electronic commerce in 1990s, the issue of privacy and security has been around. Udo [5] performed a study to investigate the concerns of online IT users to analyze the concerns in the literature. The author developed a list of concerns based on the review of the literature on the issue. According to Udo [5], following are the concerns about the use of e-mail and the Internet that were ranked in order of importance: privacy; security and threats; children protection on the Internet; e-mail safety; and censorship, impersonation and forged identity.

Business and other organizations have become aware of potential threats on the Internet and the web during last few years. However, despite increasing awareness of the associated risks, entities ranging from consumers to large businesses, still do not take advantage of available technology and processes to secure and protect their systems [6]. According to the IPTF [6], a lack of investment puts firms and consumers at greater risk, leading to financial loss to individuals, organizations and the nation as a whole. In this research, authors wanted to see if the user perception of some of the concerns regarding web and email has changed over the last few years.

These concerns have only been made very public due to the recent Facebook's data crisis scandal regarding how user data is shared with its partner Cambridge Analytica. The third party company who gained this information then tried to influence American voters during the last presidential election. A Federal Trade Commission has been in progress resulting in many other organizations revisiting their data sharing and privacy agreements. Questions about data privacy are now at the forefront due to Facebook's business model which relies on more than 1.4 billion users engaging the platform and sharing information daily and other similar organizations are starting to take notice.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Phishing case studies using a variety of methods have been performed by Aburrous, Dahal, Hossain, & Thabatah [7]. One case involved website phishing who targeted 120 employees at an organization where 52 out of the 120 targeted employees responded giving out sensitive login credentials. Even the IT department had a few employees that fell victim to this phishing website. In another case where a website survey was used 72% of those who participated made wrong determinations about the legitimacy of the website used in the experiment.

After the results were disclosed some of the employees saw the learning value of the experiments and understood the importance of the exercise; however, there were those that felt deceived and violated their privacy rights. These reactions will need to be analyzed and formal policies and procedures will need to be documented for future use adding to the challenges of the problem.

Research Method

The purpose of this study was to investigate the concerns of online users in order to ascertain what significant issues determine whether or not a user would trust an online entity. A simple 20-item questionnaire was developed and administered online to a group of volunteer students. The items were derived from the previously used questions on privacy and security issues in the literature. The survey instrument was pilot tested on some experienced online users who were also familiar with research issues on privacy and security concerns. The questions were further refined based on the comments received from the pretest, before administering it to the participants. The questions relate to simple items of concerns that an online user faces, each respondent was allowed to select one of the options from five possible choices, namely, Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree.

A second tool was used which included a phishing email with a link that emulates an authentic login portal for students in an attempt to gather information regarding users who click on the link and essentially give away their secure credentials. The link was designed in a way that users would be unaware that they will be entering vital information into a potential database that would collect their information and the results are reported and analyzed.

Data Analysis

The data collected in this study came from the 126 surveys collected from the volunteer respondents. In Udo's study, 64% of respondents agreed or strongly agreed to the statement that *Security and privacy concerns are barriers for shopping online*, while 56% of responded in the present study agreed or strongly agreed. In Udo's study, 15% of respondents agreed or strongly greed to the statement, *Feel safe*

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

when credit card information is released on the Internet. However, 23% of the respondents agreed or strongly agreed with this statement in the current study. Concerning emails, in Udo's study, about 48% agreed or strongly agreed with the statement, *Some e-mails do not come from people who appear to send them*, but in current study that value has gone up to 67%. So most of the results are similar to what they were in prior studies. But one area where there is a considerable improvement is regarding the statement, *Current encryption and passwords are sufficient or security and safety when on the Internet*, in Udo's study about 25% agreed or strongly agreed with it, but a larger percentage, 75%, agrees or strongly agrees in this study.

In this study, only three issues seem to significantly determine whether people buy online or not. The significance level of those issues is provided in Table 2. Those respondents that considered the identity of the email senders may not be correct, those who clear the browser cache cookies regularly, and those who are concerned about the people online not being who they say they are, significantly impact whether they buy online.

Table 2. T value and significance level of issues that are found to be critical in determining whether a person would buy online.

Issue	T value	Significance
Identity of email senders	3.62	.0004
Precautions, browser cache clearing	2.05	.0422
Identity of people online	3.65	.0004

The authors can attribute this to increased technical education over the years since the Udo's study; however, further analysis still shows that there is much room for improvement and that there possibly should be updates to Udo's 2001 experiment to match today's generational mindsets.

In the second experiment the phishing email link was sent to entry level college students in 5 different classes. Four of the classes were introductory and included students in all majors, whereas the last class was a junior class where most students were IT majors. The results are as follows:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Class	Students who clicked on link and submitted information	Students who reported link as a possible phishing attempt
ITEC 1001 – Introduction to Computing (Section 5)	25	3
ITEC 1001 – Introduction to Computing (Section 20)	24	4
ITEC 1001 Hybrid – Introduction to Computing (Section 7)	21	6
ITEC 1001 Hybrid – Introduction to Computing (Section 29)	24	4
ITEC 3100 – Introduction to Networks (Section 5)	6	22

From the 4 classes with introductory college students of varying majors a total of 94 out of 111 (84.7%) clicked on the phishing link and provided personal user information. In the junior IT course only 6 out of 28 (21.4%) clicked on the link. Collectively there were 100 out of 139 (71.9%) students who furnished information via the phishing link which is consistent with some findings done independent agencies at varying organizations such as the experiments reporting around the 70% mark [7].

Conclusions and Recommendations

There is no doubt that IT security has been a major concern for businesses, non-profit organizations and educational institutions for last several years. Educause [8] reports information security as the top IT issue for 2018. Performing risk analysis and developing a security strategy that can deal with emerging security threats is the key. To this end, user training, threat prediction and prevention, intrusion detection and response are elements that can go a long way in securing an organization but each organization must take steps to ensure that their employees are trained and re-trained on this on-going epidemic.

Predictions that social engineering and phishing attacks will continuously be on the rise. Billions of dollars are lost every year by corporations and internet users to these attacks so education and training is key to mitigate these security breaches. Results from the surveys and experiments show the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

importance of training and education awareness to all users. The recent use of multi-factor authentication, a method of confirming a user's claimed identify using 2 or more pieces of evidence has started to gain traction and can be used to help combat these issues. Some recommendation strategies include the following:

- Creating updated security training materials dealing with phishing, social engineering, website spoofing and other common security threats.
- Requiring mandatory training of all employees from the top down using the updated training materials.
- Employing multi-factor authentication for user accounts
- Building a community of security awareness

The authors plan to do further research on the effects of utilizing some of the recommended strategies and perhaps repeat the phishing experiment again after some time to see what effects they have. The scope and setting of the research must also be considered and perhaps the use of industry companies willing to provide experimental data can give a clearer picture of the strategies.

References

[1] Reyns, B., Henson, B., & Fisher, B. (2011). Being pursued online. *Criminal Justice and Behavior,* 38(11), 1149-1169.

[2] Gradon, K. (2013). Crime science and the Internet battlefield: securing the analog world from digital crime. *Security & Privacy*, 11(2013), 93-95.

[3] Farooq, M.U., Waseem, M., Khairi, A., and Mazhar, S. (2015) A Critical Analysis on the Security Concerns of Internet of Things (IoT). International Journal of Computer Application, 111 (7)

[4] Burak, L., Rosenthal, M., & Richardson, K. (2013). Examining attitudes, beliefs, and intentions regarding the use of exercise as punishment in physical education and sport: an application of the theory of reasoned action. *Journal of Applied Social Psychology*, 43(11), 1436-1445.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[5] Udo, G. (2001). Privacy and security concerns as major barriers for e-commerce: a survey study. Information Management & Computer Security, 9/4 [2001] 165-174

[6] Internet Policy Task Force – IPTF (2011). Cybersecurity, Innovation and the Internet Economy. The Department of Commerce. Retrieved from https://www.nist.gov/sites/default/files/documents/itl/Cybersecurity_Green-Paper_FinalVersion.pdf

[7] Aburrous, M., Dahal, K., Hossain, M., & Thabatah, F. (2010). Experiemental Case Studies for Investigating Phishing Intelligent Techniques and Attack Strategies. *Journal of Cognitive Computation*, 2(3), 242-253.

[8] Educause (2018). Educause research Snapshot. Retrieved from https://www.educause.edu/~/media/files/educause/research/2018/2018-top-ten-infographic.pdf?la=en

Table 1. Opinion of privacy and security concerns

The current security features such as encryption and passwords are sufficient to provide security and safety when on the Internet.	Strongly Agree	40	32%
	Agree	54	43%
	Neutral	21	17%
	Disagree	7	6%
	Strongly Disagree	3	2%
I feel safe when I provide my credit card information on the Internet.	Strongly Agree	8	6%
	Agree	22	17%
	Neutral	53	42%
	Disagree	33	26%
	Strongly Disagree	10	8%
Some e-mails do not come from the people that appear to send them.	Strongly Agree	29	23%
	Agree	56	44%
	Neutral	31	25%

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

	Disagree	6	5%
	Strongly Disagree	4	3%
Security and privacy concerns are barriers for my shopping online.	Strongly Agree	25	20%
	Agree	45	36%
	Neutral	34	27%
	Disagree	19	15%
	Strongly Disagree	2	2%
Do you watch for ways to control what people send you online (such as check boxes that allow you to opt-in or opt-out of certain offers)?	Strongly Agree	38	30%
	Agree	56	44%
	Neutral	22	17%
	Disagree	10	8%
	Strongly Disagree	0	0%
Do you remove browser cookies?	Strongly Agree	17	13%
	Agree	38	30%
	Neutral	43	34%
	Disagree	21	17%
	Strongly Disagree	7	6%
Do you check your computer for spy ware?	Strongly Agree	27	21%
	Agree	30	24%
	Neutral	23	18%
	Disagree	38	30%
	Strongly Disagree	8	6%
Are you concerned about online identity theft?	Strongly Agree	62	49%
	Agree	42	33%
	Neutral	14	11%
	Disagree	6	5%
	Strongly Disagree	2	2%
Are you concerned about people online not being who they say they are?	Strongly Agree	59	47%
	Agree	45	36%
	Neutral	12	10%
	Disagree	8	6%
	Strongly Disagree	2	2%
Are you concerned about people you do not know obtaining personal information about you from your online activities?	Strongly Agree	66	54%
	Agree	37	30%
	Neutral	15	12%
	Disagree	3	2%
	Strongly Disagree	2	2%

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Are you concerned that if you use your credit card to buy something on the internet your credit card number will obtained / intercepted by someone else?	Strongly Agree	51	41%
	Agree	44	35%
	Neutral	22	18%
	Disagree	7	6%
	Strongly Disagree	1	1%
Are you concerned that if you use your credit card to buy something on the internet your card will be mischarged?	Strongly Agree	42	34%
	Agree	50	40%
	Neutral	24	19%
	Disagree	7	6%
	Strongly Disagree	2	2%
Are you concerned that an email you send may be read by someone else besides the person you sent it to?	Strongly Agree	27	22%
	Agree	40	32%
	Neutral	34	27%
	Disagree	20	16%
	Strongly Disagree	4	3%
Are you concerned that an email you send someone may be inappropriately forwarded to others?	Strongly Agree	28	22%
	Agree	38	30%
	Neutral	34	27%
	Disagree	19	15%
	Strongly Disagree	7	6%
Are you concerned that an email you send someone may be printed out in a place where others could see it?	Strongly Agree	27	21%
	Agree	35	28%
	Neutral	34	27%
	Disagree	23	18%
	Strongly Disagree	7	6%
Are you concerned that a computer virus could send out emails in your name?	Strongly Agree	45	36%
	Agree	48	38%
	Neutral	19	15%
	Disagree	10	8%
	Strongly Disagree	4	3%
Are you concerned about emails you receive not being from whom they say they are?	Strongly Agree	40	32%
	Agree	51	40%
	Neutral	25	20%
	Disagree	7	6%
	Strongly Disagree	3	2%
Are you concerned that an email containing a seemingly	Strongly Agree	48	38%
	Agree	51	40%

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

legitimate internet address may be fraudulent?	Neutral	19	15%
	Disagree	8	6%
	Strongly Disagree	0	0%

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Secure and energy-efficient permutation routing protocol for wireless sensor networks deployed in space (3D)

Alain Bertrand Bomgni [#1], Garrik Brel Jagho Mdemaya [#2], Elie Tagne Fute [*3], Clementin Djamegni Tayou [#4]

[#] *Department of Mathematics And Computer Science, University of Dschang Cameroon*

[1] alain.bomgni@gmail.com, Corresponding author

[2] jaghobrel@gmail.com

[4] dtayou@gmail.com

[*] *Department of Computer Engineering,University of Buea Cameroon*

[3] eliefute@yahoo.fr

***Abstract*—A wireless sensor network is composed of several sensor nodes generally having low capacities and deployed in hostile zones. Once deployed, one node may have packets destined to another: It's the permutation routing problem. In recent years, several works have been proposed to solve this problem in single and multi-hop environments. However there are very few that incorporate security, and none has been proposed when the network is deployed in space (3D). In this paper, we propose a secure and energy-efficient permutation routing protocol for a wireless network when sensors are deployed in space. This protocol is executed in three main phases and, unlike the other protocols presented in the literature, it is secured, very energy-efficient and considers that the memory of the sensors is not infinite.**

I. Introduction

Wireless sensors networks are networks constituted of small devices massively deployed in an area in order to collect and transmit data toward one or several points, in an autonomous way [1]. Sensors composing the network generally have weak capacities of memory and energy, the access to the medium radio being the most expensive element in energy [2]. In this type of network, the security is a crucial point to study and to put forward. Indeed, wireless sensor networks have several constraints as the communication mode which is wireless: nowadays, it is very easy to read, intercept or modify data transmitted and compromise the entire network [3]. Let's add to these inconveniences the context of application of the sensors, that are generally expanded in zones where human beings can't reach; therefore it becomes necessary to develop efficient routing techniques that minimize the energizing consumption of the sensors in order to ensure the longevity of the network. In this type of network, a sensor can be deployed holding information to be transmitted to another. For example, in the military domain, a soldier may hold information that does not belong to him and that he must pass on to another soldier. Thus, in order for each soldier to carry out his mission, the latter must take possession of the various information that is intended for them [4].

The underlying problem is the permutation routing problem which is stated as follows: Consider a WSN of p stations with n items circulating in the network. Each item has a unique destination, which is one of the p stations and each sensor initially holds $\frac{n}{p}$ items. It is important to note that in general, some of the items stored in a station, say i, have not i as final destination station. The permutation routing problem is to route the items in such a way that for all i, $1 \leq i \leq p$, station i contains all its own items at final. This problem has been studied in the literature in single and multiple hop environments [5].

In this paper, we therefore study the permutation routing problem in multi-hop environment. Our objective is to propose a permutation routing protocol which is secured and energy efficient and which performs in three stages: The first stage is devoted to the clustering of the network into clusters where CHs are elected, the second stage is allocated to the routing of external items toward their directed clusters. Finally, in each clus ter, we route the internal items to their destination.

The remainder of this paper is organized as follows: In section II we present the various works dealing with the permutation routing problem; in Section III, we present our permutation routing protocol for WSNs; in section IV, we analyse security issues, then in section V we present differences between our protocol and some others protocols existing in literature, Section VI deals with some experimental results. A conclusion with open problems ends the paper.

II. Related works

The different works proposed in the literature to solve the permutation routing problem are classified into two categories.

A. Permutation routing protocols in one hop environments

Nakano and al. [6] first proposed a permutation routing protocol in a network with p stations and one communication

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

channel. Therefore, they proposed a technique of channel reservation in order to avoid collisions of messages.
In [7], A. Datta and al. presented a protocol which is more efficient than the one presented in [6] in terms of the energy consumption of the sensors and the number of wake-up slots for each of the stations in the network. However, security has not been taken into consideration in these two protocols.

B. Permutation routing protocols in multi hop environments

Many works experienced these latest years had been done in the domain of data routing and more precisely in the domain of permutation routing in a multi hop environment. Heinzelmann and al. [8] introduced a hierarchical clustering algorithm for sensor networks called LEACH which uses the random rotations of the CHs. In LEACH, The CHs send the collected information directly to the base station; which leads to a very high energy consumption. An enhancement over LEACH protocol called PEGASIS was proposed by Lindsey and al. [9]. The protocols presented in [8] and [9] are not secured and assume that the memory capacity of the sensor nodes is not limited.
Bomgni and al. in [10] have introduced a deterministic routing protocol for permutation routing in dense multi hop sensor networks which performs in 5 stages. After partitioning the network into cliques [11], this protocol first proceeds to a local distribution of the packets within each clique before making hierarchical partitioning thereafter. Lakhlef and al. proposed in [12] an improvement of the previous protocol. However, the protocols presented in [10] and [12] are not secured and assume that all network stations are awake during the entire execution of the protocols. In [13] Bomgni and al. established a new permutation routing protocol which is energy efficient and non secured. H. Lakhlef and al. [4] proposed a secure permutation routing protocol in multi-hop wireless sensor networks; but there is overhead during the clustering phase.
It should be noted that, of all the protocols presented, none has yet been proposed for a sensor network deployed in space.

III. Permutation routing protocol secured and energy efficient in a wireless sensor network

A. Assumptions and notations

1) Assumptions: We consider *p* static stations randomly deployed in space, and each station has an ID between 1 and p. The BS is situated at the center of the network, and it is the only station in which we can trust and which cannot be compromised. The network is sufficiently dense; thus each cluster will have at least one sensor.

2) Notations: Our notations are explained in table I.

TABLE I
Notations

Notations	Explanation
K_{init}	Key used to authenticate all the messages during the clustering step.
$K_{(BS,CH)}$	symetric key shared between the BS and a cluster-head.
$K_{(BS,CH*)}$	symetric key shared between the BS and all the cluster-heads.
$K_{(CH,M)}$	symetric key shared between a cluster-head and the members of his cluster.
ID_A	Identity of node A.

B. Protocol phases

After deployment, the BS computes the cryptographic parameters using an elliptic curve E(a,b,K); that is, it chooses a finite field K and an elliptic curve $y^2 = x^3 + a^2 + b$. After that, it initializes an array T containing the identities of all the sensors and a point *P* on the curve that will be used to establish a private key between nodes. The BS generates a key K_{init} and gives to all the sensors the following values: ID, K_{init}, E(a,b,K) and P.
All the messages are coding using the key K_{init} during the clustering step.

C. First stage : Secured clustering procedure

The BS partitions the network into clusters in a secured way using the same technique presented in [14].

1) Formation of crowns: The integer *l* is known by all the nodes, each node must read a string of log_2(l) bits determining the identity of its crown. The exchanged messages during the formation of crowns are secured using the K_{init} key.

2) Formation of horizontal sections: The integer *m* is known by all the nodes, each node must read a string of log_2(m) bits determining the identity of its horizontal section. Communications are secured using the K_{init} key. The value of α is given by the theorem 1.

Theorem 1: In order to permit to each sensor to communicate with his neighbours in one hop, the BS broadcasts emissions of angle α such as $\sin(\alpha) = 2/R^2{}_c$. R_c is the communication radius of a sensor.

Proof: Let us consider the figure 1; the biggest clusters are those who are far away from the BS. thus, we will use the last crowns to determine the value of α.
Let R_i be the radius of the internal sphere and R_{i+1} be the radius of the external sphere; then $4\pi R_{i+1}^2$-$4\pi R_i^2$ determines the area of the part coloured in blue. Let R_c be the communication radius of a sensor; then $4\pi R_{i+1}^2$-$4\pi R_i^2/4\pi R_c^2$ permits to split up the blue part into areas of communication of a sensor. While simplifying the precedent operation, we get $(R_{i+1}^2$-$R_i^2)/R_c^2$ **(relation 1)**

Let us consider that this area equals to the area of a isosceles trapezium where *a* is the small base, *A* is the highest base and H_2 is the height. α is the center angle. The area of this trapezium is given by (A+a)*H_2/2. Let x=a/2. Then $\sin(\alpha/2)$= x/R_i; therefore a=2*R_i*$\sin(\alpha/2)$. In the same way, A=2*R_{i+1}*$\sin(\alpha/2)$.
H_2=H-H_1 = (R_{i+1}-R_i)*$\cos(\alpha/2)$. Thus the area becomes
[(2*R_{i+1}*$\sin(\alpha/2)$ + 2*R_i*$\sin(\alpha/2)$)(R_{i+1}-R_i)*$\cos(\alpha/2)$]/2 = [2*$\sin(\alpha/2)$*$\cos(\alpha/2)$(R_{i+1}^2-R_i^2))]/2. However,

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

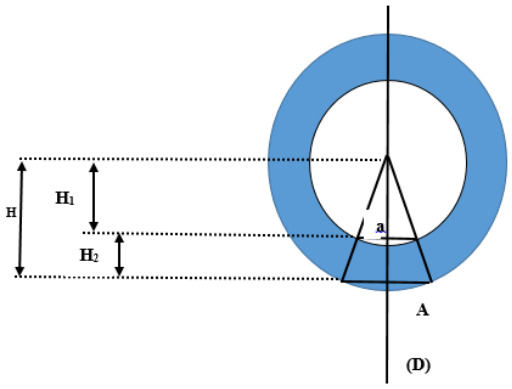


Fig. 1. Determining the value of α

2*sin(α/2)*cos(α/2) = sin(α); the new area is given by: [sin(α)*(R_{i+1}^2-R_i^2)]/2 **(relation 2)**.
While equaling relations 1 and 2, [sin(α)*(R_{i+1}^2-R_i^2)]/2 = (R_{i+1}^2-R_i^2)/R_c^2 . then we can conclude that sin(α) =2/R_c^2.

3) Formation of vertical sections: This formation is similar to the one of horizontal section, always in a secured way and the angles are computed as indicated by the theorem 1.

4) Discovering neighbours algorithm: The sensors in each cluster broadcast their coordinates while respecting the protocol CSMA/CA [15], [16] in order to avoid collisions. After reception, each sensor compares these coordinates with its own; if these are equal, it just add the transmitter in its neighbours list.

5) Cluster head election in each cluster: Since the precedent steps consumed the same capacities in each sensor, the elected cluster head is the sensor having the smallest ID in each cluster. Thereafter, as soon as the energy of the CH will be less than a threshold, the CH is re-elected and the sensor having the highest energy will be chosen. As soon as the energy of all the sensors will be less than this threshold, a new threshold is computed using the average of the energy of all the sensors in a cluster and the re-election restarts.

6) Election of two gateway nodes in each cluster: In [17] data are transmitted from the most distant cluster to the BS. Let us consider a cluster of coordinates (i,j,k); two gateway nodes are elected as follow: the first is near the cluster of coordinates (i-1,j,k) and the second is near the cluster of coordinate (i+1,j,k). Using the CSMA/CA protocol [15], sensors in each cluster perform a *ping* message to the cluster of coordinate (i-1;j;k). Sensors that receive the message register the might of the signal and send it to the CH. The sensor having the highest might is elected as gateway node in the cluster (i-1,j,k). Then it sends an acknowledgement message to the cluster (i,j,k) and the sensor receiving the message with the highest might of signal is elected as gateway node in the cluster (i,j,k). Thereafter, a gateway node will be re-elected as soon as his energy will be less than a threshold using the same technique proposed for the re-election of the CH.

7) New keys establishment: At the end of the previous stage, the BS establishes a new key with each CH ($K_{(SB,CH)}$), and the key $K_{(SB,CH*)}$ used to authenticate communication between the BS and all the CHs. Then, in each cluster, the CH establishes the keys $K_{(CH,M)}$. Finally, the key K_{init} is deleted and communications are now secured using the keys $K_{(SB,CH)}$, $K_{(SB,CH*)}$ and $K_{(CH,M)}$.
It is a key exchange in the manner of *Diffie and Hellman*, which means without communicating them directly. Each CH knows E(a,b,K) and *P*. Then each CH chooses an integer K_A and the BS chooses an integer K_B. Each CH sends to the BS the point K_AP of the elliptic curve, and the BS sends to all the CH the point K_BP. Each CH is therefore able to compute $(K_AK_B)P$; and this value is be the same for the BS. This point of the elliptic curve constitutes the secret key $K_{(SB,CH)}$. Following the same principle, the keys $K_{(SB,CH*)}$ and $K_{(CH,M)}$ are established.

D. Second stage: Routing the external items for each cluster

After the first stage, the network is divided into l×m×n clusters in which each sensor knows its neighbours. Several virtual paths relying the BS to each cluster are also created similarly to the virtual paths presented in [14]. All these paths meet at the BS thus creating a tree where the BS is the root, the nodes are the middle CHs and the leaves are the CH of the most distant clusters.

1) General Scheduling of clusters: Our goal in this section is to present the order in which the different clusters of network receive their items. To realize this, we use a similar method like the one presented by Bomgni et al. in [10]. The BS broadcasts the ordered list of ID of the different CH in the tree. After receiving, each CH identifies its position in the list and informs the members of its cluster. The notification of members in a cluster is done in parallel within the clusters and requires 1 slot. Importantly, no station is awake for more than 1 slot. Hence, this phase requires a total of *F=3l+1* slots and meanwhile all stations involved remain awake for up to *f=3* slots.

2) Identifying and Scheduling clusters That Have Items in Direction of the cluster i: Here we have to identify all the clusters having items in direction of the cluster i. To achieve this, we proceed into 4 phases.

Phase 1: sending the list of members of the cluster i to BS

Each CH knows its neighbours in its cluster; it sends this list to the BS using the key $K_{(SB,CH)}$. CH of original cluster and the BS remain awake at most 1 slot. However, the other CHs and relevant gateways remain awake during 2 slots. One slot to receive the list and the other to retransmit. Thus, each node contributing in this phase remains awake for up to 2 slots. Communication between two clusters requires a maximum of 2 slots. One to move from one station in the cluster to the gateway, another one from the gateway to the gateway of the next cluster on the path. The time used by the CH to send the list of its neighbours for the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

farthest cluster is equal to l slots. Since communication between two clusters needs at most 2 hops, then we have *2l* slots. The BS doesn't send its items; so we have (l-1) transmissions; therefore, C_1=*2(l-1)* slots are used to finish this phase, while stations are awake for at most c_1=*2(l-1)* slots.

Phase 2: Broadcasting the members list of cluster i to the other clusters

The BS broadcasts the list of the members of the cluster i. After receiving this list, the CHs record and retransmit it to theirs sons. The process is repeated until the leaves receive the list. This operation needs *5l* slots since there are *l* clusters in a path, and in each cluster the list is transmitted in 5 slots. However, the communication in the leaves needs 3 slots. Each CH is awake for 3 slots and each gateway needs to be awake for 2 slots with the exception of the BS that only transmits the list (in 1 slot) and the CH of the leaves that only receive the list (in 1 slot). Finally, this phase needs C_2=*5l-2* slots and sensors are awake for at most c_2=*7l-3* slots.

Phase 3: Identifying the items in destination to cluster i among different clusters

The goal in this phase is to determine the number of items that different clusters have in destination to cluster i. Since each item to convey is the pair (a(v),v), stations know whether the items in their possession are destined to cluster i or not. This is done using the reservation protocol presented by Nakano and al. [6]. Let $|HUB_{max}|$ be the number of stations of the cluster with maximum members. Since this phase is executed in parallel in each cluster, it will need $|HUB_{max}|$ slots. After this phase, each station knows when to transmit its items [6]. This phase needs C_3=$|HUB_{max}|$ slots while stations remain awake for c_3=*3* slots.

Phase 4: Scheduling the clusters to transfer the items to the cluster i

In the first slot, CHs in leaves's clusters wake up and send in the reserved channel of their clusters, the number of items that they have in destination to cluster i. After 3 slots, the CHs of the parents' clusters wake up and receive the numbers sent by their sons, and each of them computes the sum N_j+N_k, (N_j is the number of items that the cluster j has to transmit to cluster i, and N_k is the number of items of his sons). This sum is computed during 1 slot while all the other stations are asleep. Since different channels are used for the ascent information in the tree, it comes to the BS in *4l* slots for the most distant clusters. However, CHs in leaves just send their items (in 1 slot), each gateway node is awake for at most 2 slots and the other CHs are awake for at most 3 slots. Globally, this phase needs C_4=*4l-2* slots, and stations are awake for at most c_4=*7l-4* slots.

3) Sending External Items Identified in the Tree to the Cluster i Using the Cyclic Reception Technique: The goal now, is firstly to send all items destined to the cluster i to the BS using the key $K_{(SB,CH*)}$; then, secondly to send these items to cluster i using the key $K_{(SB,CH)}$. To achieve this, we proceed in 2 phases.

Phase 1: Broadcast of items intended to cluster i to the BS

Our job in this phase is to forward the items destined to the cluster i to the BS. To achieve this, we use the same technique presented in [13].

- **Transfer cluster i items to the cluster of upper level in the tree:** The goal here is to forward all the items intended to cluster i from the cluster in leaves to the BS. We therefore use the cyclic transmission presented in [13].
- **Cyclic reception of cluster's i item from low level clusters in the tree:** Using the same principle presented in [13], all clusters know when they will start receiving items which are intended to cluster i and arising from their low level clusters. In addition, all stations know exactly how many items they will receive from clusters of lower-level. The receipt of these items is done in accordance with channel reservation protocol [6], and using the cyclic reception technique presented in [13].

The process described above is executed until the BS receives all the items intended to cluster i. *2l* slots are necessary for an item to quit from a cluster leaf to attend the BS in the worst case. If all the items of cluster i are initially in leaves, this phase needs D_1=*2l(*$|HUB_{max}|$*)n/p* slots to finish, and the stations are awake for d_1=*2(*$|HUB_{max}|$*)n/p* slots.

Phase 2: broadcast of items intended to cluster i from the BS

After phase 1, all items intended to the cluster i are in the BS. The BS then proceeds as follows:

- **Inform the nodes about the path toward cluster i:** At the beginning of this phase, all the stations of the network are awake. The BS then broadcasts the number of items N_i intended to cluster i and the path to follow until this cluster. The message is sent through the tree until the leaves and will need *4l* slots to finish since there are l crowns and in each cluster the communication is performed in 4 slots. The leaves don't transmit the message, therefore this phase finishes after D_2=*4l-2* slots; the CH and the gateway nodes are awake for d_2=*4l-2* slots.
- **Broadcast the items to cluster i**
  all the nodes that are not on the path joining the BS and the cluster i fall asleep while the other nodes know exactly when they will wake up to receive items coming from their parents. The BS then starts the transmission.
  i) **Broadcast the items from the BS to cluster i:** The BS sends the items to his son on the path linking the BS and cluster i. Thus, in accordance with the principle of cyclic transmission [13], these items will be forwarded until cluster i.
  ii) **Cyclic reception of items by clusters located on the path to the cluster i:** The items receipt is done

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

following the principle of cyclic reception [13].

2l slots are necessary to transmit items to the most distant cluster. If the cluster i is a leave, and all his items are initially out of him, then D_3=*2l*|HUB_{max}|*n/p* slots are used to finish this phase while the station are awake for d_3=*2(*|HUB_{max}|*)n/p* slots.

E. Third Stage: Local Broadcasts in Clusters

All items in destination of a cluster are in this cluster and each station in each cluster just has to copy his items in his memory. This operation is done by applying the protocol of Nakano and al. [17], and communications are authenticated using the key $K_{(CH,M)}$. In the worst case, the cluster contains |HUB_{max}| stations. The number of items capable to be in a cluster at this stage is |HUB_{max}|n/p. Nakano and al. proved in [17] that a permutation routing in a cluster containing |HUB_{max}| and in which |HUB_{max}|n/p items are circulating can be made in *E=(2d*|HUB_{max}|*n/p)-(2(*|HUB_{max}|*-1))* slots, and stations are awake for *e=4dn/p* slots where *d=(*log_2(|HUB_{max}|*))/(*log_2*(n/p))*.

The pseudo-code of our protocol is given by the algorithm 1.

Theorem 2: Our permutation routing protocol needs *(l+m+n-3)* + |HUB_{max}| + *[18l-5+*|HUB_{max}|*[4l*$\frac{n}{p}$ *+2d*$\frac{n}{p}$ *-1]* slots to finish in the worst case and the stations are awake for *20l-3+4*$\frac{n}{p}$ *[*|HUB_{max}|*+d]* slots. with *l* the number of crowns, *m* the number of horizontal sections and *n* the number of vertical sections.

Proof:

- The formation of crowns needs (l-1) slots; for the horizontal sections (m-1) slots and for the vertical sections (n-1) slots. The clustering algorithm performs in (l+m+n-3) slots.
- The Discovering neighbours algorithm is realized in parallel in each cluster and needs |HUB_{max}| slots, each station being awake for 2 slots.
- Thereafter, the routing phase needs F+C_1+C_2+C_3+C_4+D_1+D_2+D_3+E slots to finish and stations remain awake for *f*+c_1+c_2+c_3+c_4+d_1+d_2+d_3+*e* . While simplifying these operations, we can obtain F+C_1+C_2+C_3+C_4+D_1+D_2+D_3+E = *18l-5+*|HUB_{max}|*[4l*$\frac{n}{p}$ *+2d*$\frac{n}{p}$ *-1]* slots to achieve the routing phase in the worst case, while stations remain awake for *f*+c_1+c_2+c_3+c_4+d_1+d_2+d_3+*e = 20l-5+4*$\frac{n}{p}$ *[*|HUB_{max}|*+d]* slots.

The summation of the previous values for each phase permits us to get the final result: *(l+m+n-3)* + |HUB_{max}| + *[18l-5+*|HUB_{max}|*[4l*$\frac{n}{p}$ *+2d*$\frac{n}{p}$ *-1]* slots are necessary to finish the routing phase while stations are awake for *20l-3+4*$\frac{n}{p}$ *[*|HUB_{max}|*+d]* slots.

Theorem 3: Our permutation routing protocol needs *(l+m+n-3) + 10l-3+ 2*|HUB_{max}| slots to finish in the best case and the stations are awake for *9l+3* slots.

Algorithm 1: Permutation routing protocol in a WSN deployed in 3D

Input: WSN with p stations and each station doesn't has his items.

Output: WSN with p stations and each station has all his items.

```
Begin
    First stage: Secured clustering procedure ;
    Begin
        Construction of crowns, horizontal and vertical sections ;
        Discovering neighbors ;
        Cluster head election in each cluster ;
        Election of two gateway nodes in each cluster ;
        New keys establishment ;
    End
    Second stage: Routing the external items for each cluster ;
    Begin
        General Scheduling of clusters ;
        Identifying and Scheduling clusters That Have Items in Direction of the cluster i ;
        Begin
            Phase 1: sending the list of members of the cluster i to BS ;
            Phase 2: Broadcasting the members list of cluster i to the other clusters ;
            Phase 3: Identifying the items in destination to cluster i among different clusters ;
            Phase 4: Scheduling the clusters to transfer the items to the cluster i ;
        End
        Sending External Items Identified in the Tree to the Cluster i Using the Cyclic Reception Technique ;
        Begin
            Phase 1: Broadcast of items intended to cluster i to the BS ;
            Phase 2: broadcast of items intended to cluster i from the BS ;
        End
    End
    Third Stage: Local Broadcasts in Clusters ;
End
```


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Proof: The clustering stage always needs (l+m+n-3) + $|HUB_{max}|$ slots. Thereafter, in the best case, all the stations initially have their items. Thus, during the second stage, only the following phases will be executed:

- **General Scheduling of clusters:** in *F=3l+1* slots and stations are awake during *f=3* slots;
- **sending the list of members of the cluster i to BS:** in C_1*=2(l-1)* slots and stations are awake during c_1*=2(l-1)* slots;
- **Broadcasting the members list of cluster i to the other clusters:** in C_2*=5l-2* slots, and stations are awake during c_2*=7l-3* slots;
- **Identifying the items in destination to cluster i among different clusters:** in C_3=$|HUB_{max}|$ slots while stations are awake during c_3*=3* slots.

The summation of these results establishes that *(l+m+n-3) +* $|HUB_{max}|$ *+ 3l+1+2(l-1)+5l-2+*$|HUB_{max}|$ *= ((l+m+n-3) + 10l-3+ 2*$|HUB_{max}|$ slots are necessary to achieve our protocol in the best case while the stations are awake during *3+2(l-1)+7l-3+3+2 = 9l+3* slots.

Theorem 4: Assuming that the size of each cryptographic key is the same size than n/p items, and that $|HUB_{max}|$<2$|HUB_{min}|$, then each sensor of the network has a memory capacity in O(6$\frac{n}{p}$).
Proof:

- Before the running of the protocol, it is assumed that each station has in its internal memory a total of $\frac{n}{p}$ items.
- If a sensor is a normal node, then it keeps only one key in its memory (the size of the key is in O($\frac{n}{p}$)). Besides, if this node is a CH, it will also keep 2 others keys that it shares with the BS (O($\frac{2n}{p}$)).
- During the protocol process, the worst case occurs when on the path from the BS to the cluster of maximum number of stations $|HUB_{max}|$, items have to pass through the cluster of minimum number of stations $|HUB_{min}|$. The total number of items destinate to cluster $|HUB_{max}|$ is $\frac{n}{p}$ $|HUB_{max}|$. Thus, each station of the cluster $|HUB_{min}|$ must store (n/p$|HUB_{max}|$)/$|HUB_{min}|$ items. According to our hypothesis $|HUB_{max}|$<2$|HUB_{min}|$ we obtain (n/p$|HUB_{max}|$)/$|HUB_{min}|$ < (2n/p$|HUB_{min}|$/$|HUB_{min}|$ = 2$\frac{n}{p}$. Hence, it is concluded that the memory allocated for routing items in different stations is at most 2$\frac{n}{p}$.
- To obtain the maximal memory capacity of each station in the network, we just have to sum the size of the memory for its own items $\frac{n}{p}$; the memory size for routing items 2$\frac{n}{p}$; and the memory size to store the three keys $\frac{3n}{p}$. Therefore we obtain O(6$\frac{n}{p}$).

IV. Security analysis

Our protocol uses cryptographic keys to guarantee authentication and confidentiality on the exchanged messages. Indeed, while the clustering algorithm, all the stations use a unique key. However, if an attacker wants to know this key he has to solve the discrete logarithm problem which is a NP complete problem; therefore the time spent to determine the key is enough.
Thereafter, after the clustering stage, the initial key is deleted and new keys are established. Thus, the BS establishes a new key with each CH, a new key with all the CHs for messages in broadcast and each CH establishes a key with all his cluster's members.
This method of security avoids passive attacks whose goal is to read or update data circulating in the network such as *passive attacks, injection of messages, deterioration of message, Sybil attacks and message replication*.

V. Comparative study of our protocol with some others existing protocols

In table II, we make a comparative study between our protocol and some others.

TABLE II
Comparison between our protocol and some others

Protocols	Dimension	Sensor's memory	Security management	CH management
LEACH [8]	2D	Infinite	unsecured	No use of CH
PEGASIS [9]	2D	Infinite	unsecured	Use and random rotation of CH
S. Faye and al. [18]	2D	O($\frac{2n}{p}$)	cryptography	Use of CHs without rotation
Bomgni and al. [10]	2D	infinite	Sun and al.[11]	Use of CHs without rotation
Bomgni and al. [13]	2D	O($\frac{3n}{p}$)	Sun and al.[11]	Use of CHs without rotation
Our protocol	3D	O($\frac{6n}{p}$)	ECC	Use and deterministic rotation of CH

From this table, we can conclude that:

- Our protocol is the only one implemented in dimension 3;
- Most of these protocols do not propose a CH rotation; which leads to a work overload of the latter and consequently a considerable loss of energy;
- The protocols [8], [9] and [10] assume that the memory of the sensors is infinite; which is not very realistic;
- Compared to the protocols [13], [18] and [10] in term of security, our protocol is the best. Indeed, the keys generated by our protocol can be broken only after a certain time T, and before the expiration of this time, new keys are generated.

VI. Simulation results

The presented curves are the average of 100 experiments. We made the common assumption that two nodes are neighbours if and only if their Euclidean distance is less than 1

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

km. In our implementation, the MAC layer is managed in such a way that a node can only receive one message at a time with the number of items sets to 1000. To minimize the energy consumption, we used integers less than 255; thus, the maximal size of a key is 8 bits.

A. Evolution of Sensor's Energy

The energetic model we use is similar to the one in [8]. Let ET and ER be the energy used for the transmissions and the receptions of items in the network respectively. The energetic model of Heinzelman and al. is E = ET+ER = a(e_t+e_{amp}*d^2)+ n*e_r. Each station initially has 1000J; 5J are used to transmit an item and 4j are used to receive an item. The curve in figure 2 presents the evolution of our nodes: normal nodes (in black), gateway nodes (in red) and CH nodes (in blue).
We can see that the clustering algorithm consume the same energy to all nodes; the variation of energy becomes visible while the routing stage.

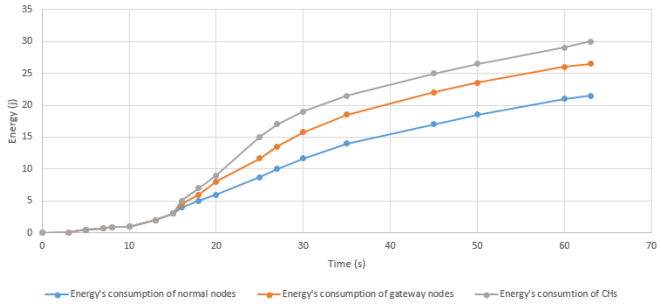


Fig. 2. Evolution of Sensor's Energy.

B. Network's lifetime

In order to put forward the efficiency of our protocol, we valued the lifetime of the network and we compared it with five other protocols: the one of Bomgni and al. [13], Sebastian Faye and al. [18], Bomgni and al. [10], Lindsey S. and al. [9] and the one of Heinzelman and al. [8]. Figure 3 shows the results.

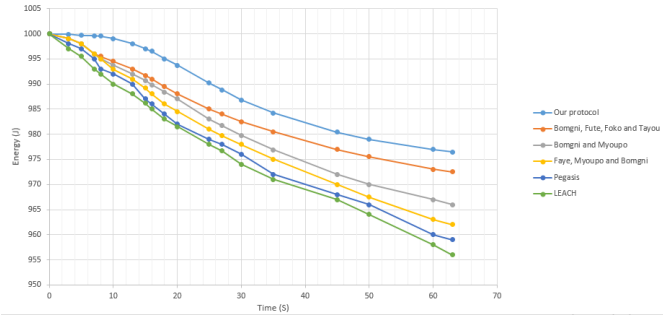


Fig. 3. Network's lifetime.

The other protocols consume more energy than our protocol because at the first stage they use the clustering algorithm of Sun and al. [11] and the one of Banerjee and al. [19] for the hierarchical clustering.

C. Awakening time of the sensors according to the duration of the simulation

The results obtained in Figure 3 can be supported by those obtained in Figure 4. Indeed, figure 4 shows the mean awakening time of the sensors as a function of the total duration of the simulation.

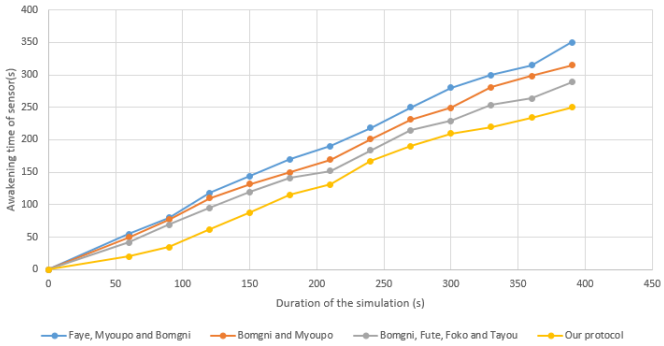


Fig. 4. Awakening time of the sensors.

As we can see, in our protocol, the sensors remain awake for a very short time compared to the protocols [13], [18] and [10]; which also causes a minimal energy loss compared to these protocols.

VII. Conclusion

We have proposed a permutation routing protocol secured and energy efficient in a WSN deployed in space. It is been process in three steps: At the first stage, we partition the network into cluster in a secured way, thereafter we route all the external items to each cluster and finally, in each cluster, each station finds its items. Contrarily to the protocols presented in [10], [14] and [12] in which the CHs are the only stations responsible of the transfer of items and the cluster's communication, in our protocol we help the CHs with the gateway nodes in order to reduce the energy's consumption. This protocol also presents an enhancement over the protocol presented in [13] which is unsecured.
However, several open problems remain. In our future work, we plan to study fault tolerance, which guarantees that normal stations receive in a finite time, the items destined to them. We also plan to apply this work on a network which is not dense and consider mobile nodes. Finally, we will ameliorate our security mechanism; this one becomes useless toward active attacks like: *blackhole attack, greyhole attack, wormhole attack.*

References

[1] Znaidi W., "Quelques propositions de solutions pour la securite des reseaux de capteurs sans fil," Ph.D. dissertation, Institut national des sciences appliquees de Lyon, 10 octobre 2010.
[2] David M. AND Guyennet H., "Etat de l'art securite dans les reseaux de capteurs," *SAR-SSI 3rd conference on Security of Network Architecture and Information Systems*, 2008.
[3] Beydoun K., "Conception d'un protocole de routage pour les reseaux de capteurs," Ph.D. dissertation, U.F.R des sciences et techniques de l'universite de FRANCHE-COMTE, 2009.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[4] Hicham L. and Myoupo J. F., "A secure permutation routing protocol in multi-hop wireless sensor networks," *international conference on security and privacy for communication networks*, 2011.
[5] Bomgni A. B., "Qualite de services dans les protocoles de multicast geographique et de routage par permutation dans les reseaux de capteurs sans fils," Ph.D. dissertation, Universite de Picardie Jules Verne, 2013.
[6] Nakano K., Olariu S. and Schwing L., "Broadcast-efficient protocols for mobile radio networks," *IEEE transactions on parallel and distributed systems, volume 10*, pp. 1276–1289, 1999.
[7] Datta A. and Zomaya A. Y., "New energy-efficient permutation routing protocol for single-hop radio network," *IEEE transactions on parallel and distributed systems, 15*, pp. 331–338, 2004.
[8] Heinzelman, Chandrakasan A. and Balakrishnan H. , "Energy-efficient communication protocol for wireless microsensor networks," *Proceedings of the 33th IEEE Hawii International Conference on systems*, pp. 3005–3014, 2000.
[9] Lindsey S. and Raghavendra C., "Pegasis: Power-efficient gathering in sensor information systems," *Proceedings of the 33th IEEE Hawii International Conference on Systems*, pp. 1125–1130, 2002.
[10] Bomgni A. B. and Myoupo J. F., "A deterministic protocol for permutation routing protocol in multi-hop wireless sensor networks," *Wireless sensor networks 2*, pp. 293–299, march 2010.
[11] Kun S., Pai P. and Peng N., "Secure distributed cluster formation in wireless sensor networks," *22nd annual Computer Security Application Conference, Las Vegas*, pp. 131–140, 2006.
[12] Lakhlef H., Bomgni A. B. and Myoupo J. F., "An efficient permutation routing protocol in multi-hop wireless sensor networks," *Proceedings of the 33th IEEE Hawii International Conference on Systems*, pp. 1125–1130, 2011.
[13] Bomgni A. B., Fute E. T., Foko M. L. and Tayou C., "A tree-based distributed permutation routing protocol in multi hop wireless sensors network," *Wireless Sensor Network*, pp. 93–105, 2016.
[14] Vianney K. T., Myoupo J. F., Fotso P. L. and Zeukeng U. K., "Virtual architecture and energy-efficient routing protocols for 3d wireless sensor networks," *International Journal of Wireless Mobile Networks vol. 9 No. 5*, October 2017.
[15] Guy P., *Les reseaux*. EYROLLES, 2006.
[16] Slimane B. J. , "Allocation conjointe des canaux de frequence et des creneaux de temps et routage avec qds dans les reseaux de capteurs sans fil denses et etendus," Ph.D. dissertation, Universite de Lorraine, 2013.
[17] Wadaa A., Olariu S., Wilson L. and M. Eltoweissy, "Training a wireless sensor networks," *Mobile Networks and Applications, volume 10*, pp. 151–168, 2005.
[18] Faye S., Myoupo J. F. and Bomgni A. B., "Heterogenous clustering for secure, energy-efficient and fault tolerant permutation routing in wireless sensor networks," *International Journal of Advanced Computer Science*, pp. 249–258, may 2013.
[19] Banerjee S. and Khuller S., "A clustering scheme for hierarchical control in multi-hop wireless networks," *Proceedings of the 20th IEEE International Conference on Computer Communications, volume 3*, pp. 1028–1037, 2001.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

An Approach for Selecting Cloud Service Adequate to Big Data

Case Study: E-health Context

Fatima Ezzahra MDARBI [1], Nadia AFIFI [1], Imane HILAL[1,2],Hicham BELHADAOUI [1]

[1] RITM Lab, EST, CED ENSEM
University Hassan II
Casablanca, Morocco

[2] Lyrica labs
Information Science School
Rabat, Morocco

Fati.mdarbi@gmail.com, Nafifi@est-uh2c.ac.ma, Ihilal@esi.com, belhadaoui_hicham@yahoo.fr

***Abstract*-The expanding Cloud computing's services offers great opportunities for consumers to find the best service and best cost. It offers a computing power and a storage space adapted especially for Big Data processing. However, it raises new challenges on how to select the best service out of the huge pool. It is time-consuming for consumers to collect the necessary information and analyze all service providers to make the right decision. Moreover, it' is a highly demanding task from a computational perspective, because the same computations may be conducted repeatedly by multiple consumers who have similar requirements. Therefore, in this paper, we propose an approach based on Analytic Hierarchy Process (AHP) method, which manages the selection of the Cloud Service adequate to Big Data based on its parameters and criteria. We applied this approach on a case study in order to validate its efficity. The studied case is about the selection of the adequate Cloud Service for Big Data in the context of National Health Service (NHS) of United Kingdom (UK).**

***Keywords*: Cloud Service; PAAS; IAAS; SAAS; Big Data; MCDM; AHP; E-health.**

I. INTRODUCTION

Organizations are unable to manage, manipulate, process, share, retrieve, and analyze the Big Data using traditional software tools. Those latter are generally costly and time-consuming during data processing. Big Data is a collection of complex data with massive volume; it needs tools for a real time data management and analysis capabilities. Whereas Cloud computing is the distributed computing model, it is a trending solution for analyzing Big Data, and providing both computing facilities and resources for users. The aim of the Cloud model is to increase the opportunities for Cloud users by accessing leased infrastructure and software applications from anywhere anytime and for any data. However, the strong growth of Cloud Services makes it difficult for potential users to decide which options are best suited to their needs. Companies need to take a rational approach to ensure they choose the most appropriate Cloud Service for their Big Data.

Many approaches have dealt with the problem of Cloud Service selection. Among these approaches, we identify AHP. AHP is a Multi-Criteria Decision Making (MCDM) method that addresses decision problems. It highlights the mutual influence of criteria, which is represented both quantitatively and qualitatively. The original use of AHP was to rank a limited number of alternatives for a limited number of criteria.

The problem of selecting Cloud Service can be seen as a MCDM case, where decision makers should select from among a set of alternatives those that best fit their criteria. These criteria are usually conflicting, and each has its importance in the decision making process.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

This paper is organized as follows: In Section 2 we introduce the Cloud computing services, Big Data criteria and Cloud Service criteria. Section 3 presents AHP and related work. Section 4 details our proposal based on the application of the AHP in order to select the adequate Cloud Service for Big Data. Then we end up with conclusion and some future works.

II. Context of Cloud Service and Big Data

A. Cloud Computing Services

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, applications, storage, and services). It can be quickly provisioned and released with minimal management effort or interaction with the service provider. This Cloud model has five main characteristics: (i) On-demand self-service, (ii) Broad network access, (iii) Resource pooling, (iv) Rapid elasticity, and (v) Measured service. It offers three service models: (i) Software as a Service (SaaS), (ii) Platform as a Service (PaaS), and (iii) Infrastructure as a Service (IaaS). And for deployment, it proposes four models : (i) Private Cloud, (ii) Community Cloud, (iii) Public Cloud, and (iv) Hybrid Cloud [1].
Cloud computing offers software dematerialization services [2] as shown in Fig. 1:

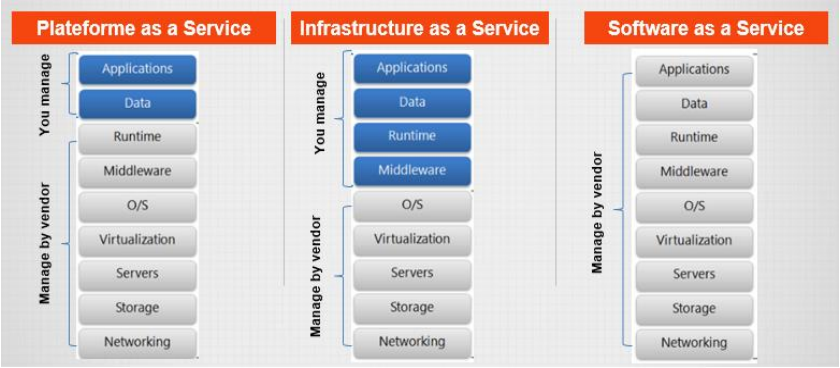


Figure 1. Cloud Services

- PaaS: This service model provides to the consumer the capability to deploy on Cloud infrastructure [3].
- IaaS: This service model provides access to a virtualized IT infrastructure. Virtual machines on which the consumer can install an operating system and applications are made available [4].
- SaaS: In this service model, applications are made available to consumers. Applications can be manipulated using a web browser [5].

B. Big Data Criteria

Big Data refers to the flood of digital data coming from many digital sources, including sensors, digitizers, scanners, numerical modelling, mobile phones, Internet, videos, e-mails and social network [6]. The main characteristics of Big Data are the five V's: Volume, Velocity, Variety, Value and Veracity [7] as shown in Fig. 2. These Big Data characteristics are the criteria used in our proposed approach.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

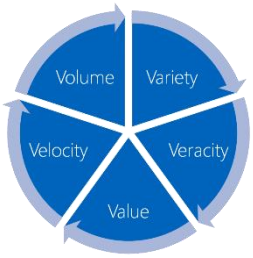


Figure 2. Five V's of Big Data

• Volume (Vol): represents the terabytes of Data produced daily by many platforms such as social networks. This amount of data is definitely diffi-cult to be handled using the existing traditional systems [6].

• Velocity (Vel): Data velocity measures the speed of data creation, streaming, and aggregation. E-Commerce has rapidly increased the speed and rich-ness of data used for different business transactions [7].

• Variety (Var): Data being produced is not of single category as it includes not only the traditional data, but also the semi-structured data from various resources like Web Pages, Web Log Files, social media sites, e-mail, documents, and sensor [6]..

• Value (Val): Is the most important aspect of Big Data. Given that the potential value of Big Data is enormous, it remains useless until its value is discovered. Moreover, since the Big Data's IT infrastructure is costly, its exploitation and value must justify the ROI for investors [8].

• Veracity (Ver): Includes trust and uncertainty issues of the analyzed data and its out coming [9].

C. Cloud Service Criteria

Many factors are involved in selection of a Cloud Service. Based on literature reviews, [10] proposed the most common criteria to choose the right Cloud Service namely: Functionality, Vendor Reputation, and Cost. [11] emphasized the importance of the Architecture and Usability, in addition [12] highlighted the Performance factor.

• Functionality (Fun): The functionality is the ability or state of being functional [13]. The need for functionality is obvious; since users will select the right system that provides suitable functions adequate to their needs [14].

• Architecture (Arc): The architecture factors are: (i) The ability to integrate with other applications; (ii) The ability to maintain reasonable response time for users even during peak load; (iii) The ability to remain available for the users for given time windows. It requires vendors to deploy monitoring and diagnostic tools; (iv) Security is considered to be the major concern [11].

• Performance (Per): The performance is determined by the properties of system's constituent parts (e.g. sensors, signal processing, and pattern recognition engine) [13].

• Usability (Usa): Usability includes (i) user interface, (ii) facets such as intuitiveness, (iii) ease-of-use for frequently required tasks and aesthetic nature of graphical elements, (iv) availability of easy-to-use user manuals, (v) eLearning modules, (vi) context-sensitive help, and (vii) offline support that let users work on system in offline mode and let them synchronize once connected to internet [13]. As cited in [14] both usability and functionality are task related and also people related.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

• Vendor reputation (Ven): Vendor reputation involves customer perceptions of the vendor's public image, innovativeness, quality of product and service, and commitment to customer satisfaction [14]. Customers can determine vendor reputation based on an evaluation of the vendor's past performance and behavior. Reputation is associated with brand equity and firm credibility; it is also viewed as a sign of trustworthiness [15].

• Cost (Cos): Cost includes one-time implementation's costs and annual subscription. Usually, cost of initial consulting, configuration efforts, etc. is covered under one-time implementation, while cost of hardware and support personnel is covered under annual subscription [13].

The cited criteria in sub section 1 and 2 of the context are the most popular for both Big Data and Cloud services. Such it is difficult to list and use an exhaustive criteria's list; we propose to focus on those already cited. However, our proposed approach is not specified to only those criteria, it could be generalized for several criteria.

III. Analytical Hierarchy Process

A. Presentation

AHP is a method for ranking decision alternatives and selecting the best one when the decision maker has multiple criteria. It answers the question, "Which one?". By using AHP, the decision maker selects the best alternative that meets his criteria. He develops a numerical score to rank each alternative based on how well it fits best his needs [16].

The AHP is a powerful, flexible and widely used method for complex problems, which consider the numeric scale for the measurement of quantitative and qualitative performances in a hierarchical structure. This is a value approach to the pairwise comparisons. It is one of the few MCDM approaches capable of handling many criteria. The most important characteristic of the AHP is combining knowledge, experience, and individual opinions in a logical way [17]. AHP was first proposed by Saaty [18], and it is one of the most commonly used methods for solving MCDM problems in different fields [19], such as:
Economic/Management problems [16], Technological problems [20], Political problems [21], Social problems [22], Big Data [17], Cloud Computing [23], [24].

Using AHP, expert opinions and evaluation can be designed in a simple hierarchy system with levels, from the highest to the lowest. The application of AHP to a complex problem involves six essential steps [21]:

• Step1: Define the unstructured problem, and state clearly the objectives and outcomes.

• Step2: Decompose the complex problem into a hierarchical structure with decision elements (criteria and alternatives) as shown in Fig 3.

• Step3: Employ pairwise comparisons among decision elements in order to refill criteria's comparison matrix using the scales presented in Table 1.

• Step4: Repeat step 3 to form the pairwise matrix of alternatives for each criterion and calculate the final vector for each matrix.

• Step5: Check the consistency property of matrices to ensure that the judgments of decision makers are rational.

• Step6: Aggregate the relative weights of decision elements to obtain an overall rating of the alternatives, using the vectors obtained from Step 3 and Step 4.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

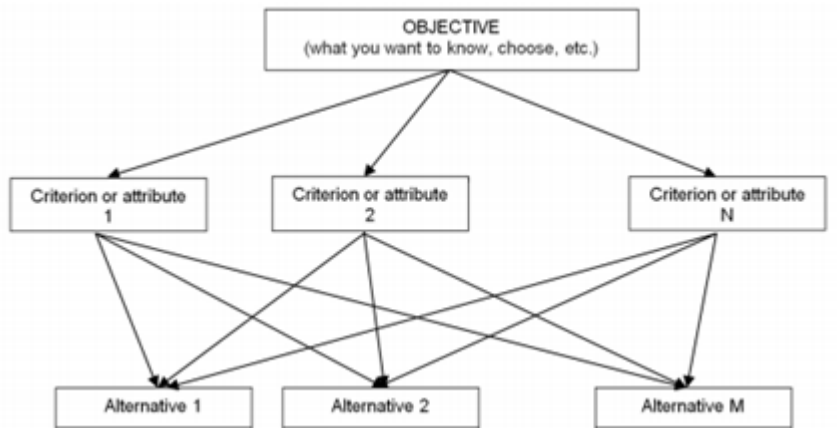


Figure 3. AHP Hierarchy

To make the comparisons cited in the steps above, we need a scale of numbers that indicates how much one element (criterion or alternative) is more important or dominant than another. Table 1 shows the used scale [25].

TABLE 1
PAIRWISE COMPARISON SCALE

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
2	Weak or slight	
3	Moderate importance	Experience and judgement slighty favour one activity over another
4	Moderate plus	
5	Strong importance	Experience and judgement slighty favour one activity over another
6	Strong plus	
7	Very strong or demonstrated importance	An activity is favoured one activity over another; its dominance demonstrated in practice
8	Very, very strong	
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation

B. Related Work

AHP has been used in many works dealing with the selection of Cloud Service. [26] presented a consumer-centered Cloud Service selection based on AHP. They adopted five criteria, namely: (i) response time, (ii) throughput, (iii) availability, (iv) reliability, and (v) cost. Weights are assigned to each one of these criteria to exhibit there importance in the decision making process. To validate their proposals, they performed tests on three types of Cloud services.

Work presented in [27] concerns an AHP-based methodology for selecting a Cloud Service for companies that would like to reduce the cost of using the software. The main contribution of this work is defining other costs beside

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

the financial one; such as costs related to scalability and risk. This is done by calculating the gain of scalability, and this by improving the agility obtained through Cloud Services. In [28], authors presented a mathematical decision model for selecting Cloud Services based on Linear Optimization and using AHP.

[29] proposed an AHP-based framework for QoS assurance to provide Cloud Services that meets both user and application requirements in terms of Service Level Agreements. While, [11] suggested an AHP-based approach for the selection of SaaS Cloud Services. [27] Presented a Cloud Service evaluation index system based on AHP. They used four evaluation criteria: (i) cost, (ii) security, (iii) reputation, and (iv) QoS. AHP is used to determine the weights of each criteria and to score nominated Cloud Services

[17] Provide an overview of Big Data analytics platforms. They proposed an AHP model, which offers a significant evaluation method that help private and public institutions selecting the suitable Big Data analytics platforms.

To summarize, we can point out that AHP is a tool that has found its use in many areas, especially in the case of Cloud Service selection. The success of the method is a consequence of its simplicity and robustness. Despite all the work cited below based on AHP, there is a lack of its use for selecting the right Cloud Service for Big Data processing.

IV. Our Proposal Approach

Based on both robustness and simplicity of the AHP, we conducted our works to select the Cloud Service adequate to Big Data. We applied this approach on a case study in order to validate its efficity. The studied case is about the selection of the adequate Cloud Service for Big Data in the context of National Health Service (NHS) of United Kingdom (UK).

NHS is the organization responsible for all healthcare services of UK. Its main concern is to manage the entire medical field characterized by massive, critical and heterogeneous Data. In recent years, the volume of data available within the NHS has increased exponentially. Modern day computing power, combined with the drop in price of data storage, means Big Data analytics is becoming more achievable [30].

To deal with the particularity of medical Data, NHS has a great potential to achieve more using the Cloud Services [31] to store and analyze their Data. This will induct a reduction in storage costs and it offers an open pathway to Big Data analytics.

However, NHS is facing the problem of moving its data in the Cloud environment. To deal with this issue, we propose to apply our approach to assist NHS select the most adequate Cloud Service.

First, we start with hypothesis to apply the AHP to our case study. We suppose that:

1. We focused on three kinds of Cloud Service models (SAAS, PASS, and IAAS).
2. We based on criteria mentioned in section 2: functionality, usability, architecture, vendor, performance, cost, volume, velocity, variety, value and veracity

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE 2
BIG DATA CHARACTERISTICS OF NHS [32]

5Vs of Big Data	NHS's Data Characteristics
Volume	Increasing size
Velocity	Speed of generation and processing
Variety	Heterogeneous Formats (XML, CSV, Multimedia, etc.)
Veracity	Critical authenticity
Value	Significant

As mentioned in AHP's section, we have 6 steps to be applied to the studied case:

Step1: Our issue is to select a Cloud Service adequate to NHS's Big Data. In or-der to achieve this goal, multiple criteria to select a Cloud Service were determined in Section 2 and then compared according to their importance. Finally, the most appropriate Cloud Service will be selected according to the predetermined criteria.

Step2: As it is shown in Fig. 4, the hierarchy structure of our announced problem is composed from 3 layers. First layer is the one that defines the main goal of the problem: "appropriate Cloud Service for NHS's Big Data". Layer 2 defines the selection's criterion. Layer 3 covers alternatives of the Cloud Services (SAAS, PAAS, IAAS).

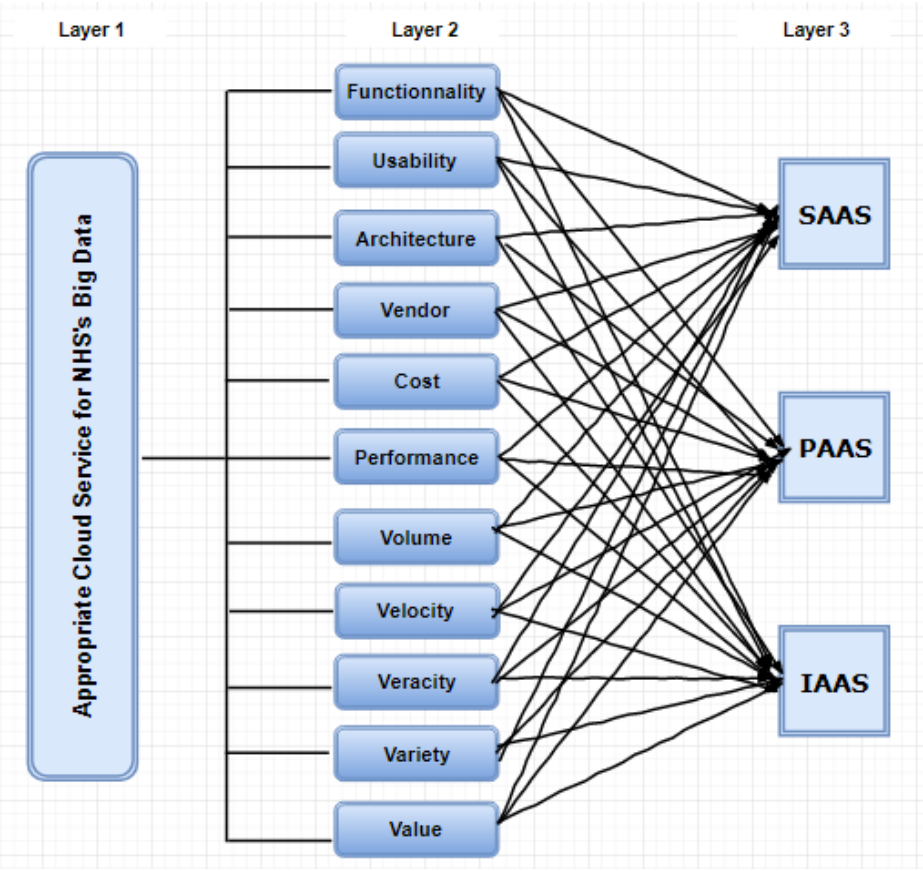


Figure 4. Hierarchy structure of AHP's step 2

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Step3: We have constructed a pairwise comparison matrix, as shown in Table 3. This matrix is composed of 2 parts: a first part (in blue) concerns the comparison of the criteria and the second (in green) included the calculated results based on data of the first part of the matrix .

TABLE 3
Pairwise Matrix

	Fun	Usa	Arc	Ven	Cos	Per	Vol	Vel	Var	Val	Ver	%Fun	% Usa	%Arc	% Ven	% Cos	% Per	% Vol	% Vel	% Var	% Val	% Ver	Sum%VCRI	Sum%VCRI/N
Fun	1	5	7	3	3	5	1	5	3	3	3	0,227	0,138	0,632	0,077	0,109	0,262	0,207	0,322	0,2217	0,083	0,093	2,371	0,216
Usa	1/5	1	1/7	5	3	1/7	1/7	1/3	1/5	3	3	0,045	0,028	0,013	0,128	0,109	0,007	0,030	0,021	0,0148	0,083	0,093	0,572	0,052
Arc	1/7	7	1	5	3	5	1	3	3	5	3	0,032	0,193	0,090	0,128	0,109	0,262	0,207	0,193	0,2217	0,138	0,093	1,668	0,152
Ven	1/3	1/5	1/5	1	1	1/3	1/7	1/5	1/3	1/3	1/3	0,076	0,006	0,018	0,026	0,036	0,017	0,030	0,013	0,0246	0,009	0,010	0,265	0,024
Cos	1/3	1/3	1/3	1	1	1/3	1/7	1/3	1/3	5	5	0,076	0,009	0,030	0,026	0,036	0,017	0,030	0,021	0,0246	0,138	0,155	0,562	0,051
Per	1/5	7	1/5	3	3	1	1/5	1	1	7	7	0,045	0,193	0,018	0,077	0,109	0,052	0,041	0,064	0,0739	0,193	0,216	1,084	0,099
Vol	1	7	1	7	7	5	1	3	3	5	3	0,227	0,193	0,090	0,179	0,255	0,262	0,207	0,193	0,2217	0,138	0,093	2,060	0,187
Vel	1/5	3	1/3	5	3	1	1/3	1	1	3	3	0,045	0,083	0,030	0,128	0,109	0,052	0,069	0,064	0,0739	0,083	0,093	0,831	0,076
Var	1/3	5	1/3	3	3	1	1/3	1	1	3	3	0,076	0,138	0,030	0,077	0,109	0,052	0,069	0,064	0,0739	0,083	0,093	0,865	0,079
Val	1/3	1/3	1/5	3	1/5	1/7	1/5	1/3	1/3	1	1	0,076	0,009	0,018	0,077	0,007	0,007	0,041	0,021	0,0246	0,028	0,031	0,341	0,031
Ver	1/3	1/3	1/3	3	1/5	1/7	1/3	1/3	1/3	1	1	0,076	0,009	0,030	0,077	0,007	0,007	0,069	0,021	0,0246	0,028	0,031	0,380	0,035
SumVCRI	4,410	36,200	11,076	39,000	27,400	19,095	4,829	15,533	13,533	36,333	32,333	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	11,000	1,000

Part I Part II

The criteria are compared with each other while respecting the initial objective. We have assigned a value between 1 and 9 as specified in Table 1, this value indicates the degree of importance of one criterion with respect to the other based on the needs of the NHS [32] Table 4 shows an example of choosing some value when comparing two criteria. We used these values to fill the part I of the matrix in Table 3.

TABLE 4
Example of pairwise comparison scales for the criterion Functionality

Intensity of importance	Definition	Explanation
1	Functionality/ volume	Functionality and volume contribute equally to the goal
5	Functionality/velocity	Functionality has a strong importance than velocity
1/5	Velocity/functionality	Inverse of functionaliy/velocity

For the part II of the matrix, we set CRI the criteria: CRIi ϵ {Fun, Usa, Arc, Ven, Cos, Per, Vol, Vel, Var, Val, Ver} and VCRIij the pairwise comparison value of the criterion i and the criterion j, where i,j ϵ {1,2,….,N} N number of criteria. We Set SumVCRIj as the sum of column j's elements.

$$SumVCRI_j = \sum_{i=1}^{N} VCRIij. \quad (1)$$

%VCRIij represent the percentage of each pairwise comparison value VCRIij.

$$\%VCRIij = VCRIij / SumVCRIj \quad (2)$$

Sum%VCRIi is the sum of all the %VCRIij of the line i.

$$Sum\%VCRI_i = \sum_{j=1}^{N} \%VCRIij \quad (3)$$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Finally, Sum%VCRIi must be divided by N for each line i in order to obtain the fi-nal vector to be used in step 6.

Step 4: According to NHS characteristics (Connecting for Health,2011) and for every criterion CRIi we repeat step 3 to form the pairwise matrix of alternatives ALT for each criterion CRI and calculate the final vector for each matrix as shown in Table 5 to 15. We set:

1) ALT the alternatives: ALTi ϵ {SAAS, PAAS, and IAAS}

2) VALTij the pairwise comparison value of the alternative i and the alternative j

3) SumVALTj as the sum of column j's elements.

i,j ϵ {1,2,….,M} M number of alternatives.

$$SumVALT_j = \sum_{i=1}^{M} VALTij \quad (4)$$

%VALTij represent the percentage of each pairwise comparaison value VALTij.

$$\%VALT_{ij} = VALTij / SumVALTj \quad (5)$$

$Sum\%VALT_i$ is the sum of all the $\%ALT_{ij}$ of the line i.

$$Sum\%VALT_i = \sum_{j=1}^{M} \%VALTij \quad (6)$$

TABLE 5
Pairwise comparison of alternatives for functionality

Fun	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	5	7	0,745	0,714	0,778	2,237	0,847
PAAS	1/5	1	1	0,149	0,143	0,111	0,403	0,153
IAAS	1/7	1	1	0,106	0,143	0,111	0,360	0,137
SumVALT	1,343	7,000	9,000	1,000	1,000	0,889	2,640	1,000

TABLE 6
Pairwise comparison of alternatives for usability

Usa	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	5	7	0,745	0,789	0,636	2,171	0,724
PAAS	1/5	1	3	0,149	0,158	0,273	0,580	0,193
IAAS	1/7	1/3	1	0,106	0,053	0,091	0,250	0,083
SumVALT	1,343	6,333	11,000	1,000	1,000	1,000	3,000	1,000

TABLE 7
Pairwise comparison of alternatives for vendor reputation

Ven	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	1	1	0,333	0,333	0,333	1,000	0,333
PAAS	1	1	1	0,333	0,333	0,333	1,000	0,333
IAAS	1	1	1	0,333	0,333	0,333	1,000	0,333
SumVALT	3,000	3,000	3,000	1,000	1,000	1,000	3,000	1,000

TABLE 8
Pairwise comparison of alternatives for architecture

Arc	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	1/3	1/3	0,143	0,077	0,200	0,420	0,140
PAAS	3	1	1/3	0,429	0,231	0,200	0,859	0,286
IAAS	3	3	1	0,429	0,692	0,600	1,721	0,574
SumVALT	7,000	4,333	1,667	1,000	1,000	1,000	3,000	1,000

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE 9
Pairwise comparison of alternatives for cost

Cos	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	5	7	0,745	0,789	0,636	2,171	0,724
PAAS	1/5	1	3	0,149	0,158	0,273	0,580	0,193
IAAS	1/7	1/3	1	0,106	0,053	0,091	0,250	0,083
SumVALT	1,343	6,333	11,000	1,000	1,000	1,000	3,000	1,000

TABLE 10
Pairwise comparison of alternatives for performance

Per	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	5	3	0,652	0,714	0,600	1,966	0,655
PAAS	1/5	1	1	0,130	0,143	0,200	0,473	0,158
IAAS	1/3	1	1	0,217	0,143	0,200	0,560	0,187
SumVALT	1,533	7,000	5,000	1,000	1,000	1,000	3,000	1,000

TABLE 11
Pairwise comparison of alternatives for volume

Vol	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	2	2	0,500	0,571	0,400	1,471	0,490
PAAS	1/2	1	2	0,250	0,286	0,400	0,936	0,312
IAAS	1/2	1/2	1	0,250	0,143	0,200	0,593	0,198
SumVALT	2,000	3,500	5,000	1,000	1,000	1,000	3,000	1,000

TABLE 12
Pairwise comparison of alternatives for velocity

Vel	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	3	3	0,600	0,600	0,600	1,800	0,600
PAAS	1/3	1	1	0,200	0,200	0,200	0,600	0,200
IAAS	1/3	1	1	0,200	0,200	0,200	0,600	0,200
SumVALT	1,667	5,000	5,000	1,000	1,000	1,000	3,000	1,000

TABLE 13
Pairwise comparison of alternatives for veracity

Ver	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	5	3	0,652	0,789	0,429	1,870	0,623
PAAS	1/5	1	3	0,130	0,158	0,429	0,717	0,239
IAAS	1/3	1/3	1	0,217	0,053	0,143	0,413	0,138
SumVALT	1,533	6,333	7,000	1,000	1,000	1,000	3,000	1,000

TABLE 14
Pairwise comparison of alternatives for variety

Var	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	1	1	0,333	0,333	0,333	1,000	0,333
PAAS	1	1	1	0,333	0,333	0,333	1,000	0,333
IAAS	1	1	1	0,333	0,333	0,333	1,000	0,333
SumVALT	3,000	3,000	3,000	1,000	1,000	1,000	3,000	1,000

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE 15
Pairwise comparison of alternatives for value

Val	SAAS	PAAS	IAAS	% SAAS	% PAAS	%IAAS	Sum%VALT	Sum%VALT/M
SAAS	1	5	3	0,652	0,789	0,429	1,870	0,623
PAAS	1/5	1	3	0,130	0,158	0,429	0,717	0,239
IAAS	1/3	1/3	1	0,217	0,053	0,143	0,413	0,138
SumVALT	1,533	6,333	7,000	1,000	1,000	1,000	3,000	1,000

In Table 5 to15, Sum%VALT must be divided by M for each line i in order to obtain the final vector to be used in step 6 . It represent the averages of each alternative ALT for each criteria CRI.

Step5: we checked the consistency property of matrices, in fact as it shown in previous tables the obtained results are homogeneous.

Step 6: We used vector Sum%VCRI/N obtained in step 3 as shown in Table 4 to fill the first line of Table 16

For each criterion CRI we used the vectors Sum%VALT/M obtained in step 4 as shown in Tables 5 to15 in order to fill columns 1 to N in Table 16.

we calculate Result according to the following formula.

$$\text{Result}_i = \sum_{j=1}^{N} \frac{\text{Sum\%VCRI}_1}{N} * \frac{\text{Sum\%VALT}_1}{M} \quad (7)$$

TABLE 16
Synthesis results

	Fun	Usa	Arc	Ven	Cos	Per	Vol	Vel	Var	Val	Ver	Result
Sum%VCRI/N	0,216	0,052	0,152	0,024	0,051	0,099	0,191	0,076	0,079	0,031	0,035	
Sum%VALT/M (SAAS)	0,847	0,724	0,140	0,333	0,724	0,655	0,490	0,600	0,333	0,623	0,623	0,557
Sum%VALT/M (PAAS)	0,153	0,193	0,286	0,333	0,193	0,158	0,312	0,200	0,333	0,239	0,239	0,236
Sum%VALT/M (IAAS)	0,137	0,083	0,574	0,333	0,083	0,187	0,198	0,200	0,333	0,138	0,138	0,240

Result and synthesis:

Following the AHP methodology, paired comparisons of criteria, and the com-parison for alternatives by criterion was made according to their intensity of im-portance and respecting a fundamental scale of absolute numbers. Then, all the calculations were performed to find normalized values for each alternative. The final decision is taken based on these normalized values.

Table 16 shows the final weights for the selected alternatives for NHS use case. Based on requirement defined in table 2, the normalized values obtained for each alternative are: 0.557 for SAAS, 0.240 for IAAS and 0.236 for PAAS. We can summarize that SAAS is the most qualified Cloud Service for NHS case.

VI. Conclusion and Perspectives

In this paper, we provide an overview of the use of AHP in the field of cloud service, and Big Data platforms. We highlight the use of the method for selecting an appropriate cloud service for Big Data. This overview allowed us to deduce that there is a lack of AHP's use in this area.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

We developed an approach that offers a simple and efficient evaluation method that can help organizations to select the most suitable cloud service ade-quate to their Big Data. To address this issue we based on AHP. In order to pro-vide complete understanding of the process we apply our approach to NHS case study.

The selection of best cloud service satisfying most of the requirements among the available alternatives is a MCDM problem. The AHP process involves multi-ple criteria and alternatives. We begin by developing a hierarchy model of 3 lay-ers: (i) goal, (ii) criteria, and (iii) alternatives. We first compare criteria pairwise with respect to the desired goal. Then we compare the alternatives pairwise for each criterion.

We calculate the relative weights of decision elements to obtain synthesis results. Finaly, we can make a final decision based on these synthesis results.

In our future works, we intend to automate our approach in order to facilitate its use, reduce calculations rate and obtain the result in few times.

The weight assignement strategy will be difficult in the case where number of cri-teria is huge. To overcome this difficulty, we will use fuzzy logic to affect weights.

We will conduct a comparative study of different Cloud Service selection meth-ods to determine the strengths and weaknesses of each.

We also plan to treat the dependability of Big Data in cloud environment in order to determine the attitude of Cloud systems to complete the features required by Big Data.

REFERENCES

[1] S. Lee, H. Park, et Y. Shin, « Cloud Computing Availability: Multi-clouds for Big Data Service », in *Convergence and Hybrid Information Technology*, 2012, p. 799-806.
[2] Q. Zhang, L. Cheng, et R. Boutaba, « Cloud computing: state-of-the-art and research challenges », *J Internet Serv Appl*, vol. 1, nº 1, p. 7-18, mai 2010.
[3] T. Grance et P. Mell, « The NIST Definition of Cloud Computing », 25-oct-2011.
[4] Q. MACHU, « Un datacenter dans le cloud Public/HA IaaS », ÉCOLE POLYTECHNIQUE DE L'UNIVERSITE FRANÇOIS RABELAIS DE TOURS, mai 2015.
[5] M. Mimoune, « Etude sur la sécurité du cloud computing ».
[6] A. Katal, M. Wazid, et R. H. Goudar, « Big data: Issues, challenges, tools and Good practices », in *2013 Sixth International Conference on Contemporary Computing (IC3)*, 2013, p. 404-409.
[7] S. Kaisler, F. Armour, J. A. Espinosa, et W. Money, « Big Data: Issues and Challenges Moving Forward », in *2013 46th Hawaii International Conference on System Sciences*, 2013, p. 995-1004.
[8] Ishwarappa et J. Anuradha, « A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology », *Procedia Computer Science*, vol. 48, p. 319-324, janv. 2015.
[9] J. S. Ward et A. Barker, « Undefined By Data: A Survey of Big Data Definitions », *arXiv:1309.5821 [cs]*, sept. 2013.
[10] J. S. Valacich, J. F. George, et J. A. Hoffer, *Essentials of systems analysis and design*, 5th ed. Upper Saddle River, N.J Prentice Hall, 2011.
[11] M. Godse et S. Mulik, « An Approach for Selecting Software-as-a-Service (SaaS) Product », in *2009 IEEE International Conference on Cloud Computing*, 2009, p. 155-158.
[12] J. O'Brien et G. Marakas, *Introduction to Information Systems*, 13 edition. Boston: McGraw-Hill/Irwin, 2005.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[13] J. W. Gardnert et P. N. Bartlett, « Performance Definition And Standardisation Of Electronic Noses », in *The 8th International Conference on Solid-State Sensors and Actuators, 1995 and Eurosensors IX.. Transducers '95*, 1995, vol. 1, p. 671-674.
[14] N. C. Goodwin, « Functionality and Usability », *Commun. ACM*, vol. 30, nº 3, p. 229–233, mars 1987.
[15] Y. Zhang, Y. Fang, K.-K. Wei, E. Ramsey, P. McCole, et H. Chen, « Repurchase intention in B2C e-commerce—A relationship quality perspective », *Information & Management*, vol. 48, nº 6, p. 192-200, août 2011.
[16] A. Özdağoğlu et G. Özdağoğlu, « Comparison of AHP and fuzzy AHP for the multi-criteria decision making processes with linguistic evaluations », *Sözel değerlendirmeli çok kriterli karar verme süreçleri için AHS ve bulanık AHS yöntemlerinin karşılaştırılması*, 2007.
[17] M. Lněnička, « AHP Model for the Big Data Analytics Platform Selection », *Acta Informatica Pragensia*, vol. 4, nº 2, p. 108-121, déc. 2015.
[18] T. L. Saaty, « Group Decision Making and the AHP », in *The Analytic Hierarchy Process*, Springer, Berlin, Heidelberg, 1989, p. 59-67.
[19] L. G. Vargas, « An overview of the analytic hierarchy process and its applications », *European Journal of Operational Research*, vol. 48, nº 1, p. 2-8, sept. 1990.
[20] A. Arbel, « Venturing into new technological markets », *Mathematical Modelling*, vol. 9, nº 3, p. 299-308, janv. 1987.
[21] A. Arbel et L. G. Vargas, « Preference simulation and preference programming: robustness issues in priority derivation », *European Journal of Operational Research*, vol. 69, nº 2, p. 200-209, sept. 1993.
[22] P. T. Harker et L. G. Vargas, « The Theory of Ratio Scale Estimation: Saaty's Analytic Hierarchy Process », *Management Science*, vol. 33, nº 11, p. 1383-1403, nov. 1987.
[23] B. R. Meesariganda et A. Ishizaka, « Mapping verbal AHP scale to numerical scale for cloud computing strategy selection », *Applied Soft Computing*, vol. 53, nº Supplement C, p. 111-118, avr. 2017.
[24] S. C. Nayak et C. Tripathy, « Deadline sensitive lease scheduling in cloud computing environment using AHP », *Journal of King Saud University - Computer and Information Sciences*, juin 2016.
[25] T. L. Saaty, « Decision making with the analytic hierarchy process », *International Journal of Services Sciences*, vol. 1, nº 1, p. 83-98, janv. 2008.
[26] M. Sun, T. Zang, X. Xu, et R. Wang, « Consumer-Centered Cloud Services Selection Using AHP », in *2013 International Conference on Service Sciences (ICSS)*, 2013, p. 1-6.
[27] G. Nie, Q. She, et D. Chen, « Evaluation Index System of Cloud Service and the Purchase Decision- Making Process Based on AHP », *Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE2011)*, vol. 112, p. 345-352, janv. 2011.
[28] B. Martens et F. Teuteberg, « Decision-making in cloud computing environments: A cost and risk based approach », *Inf Syst Front*, vol. 14, nº 4, p. 871-893, sept. 2012.
[29] S. Khaddaj, « Cloud Computing: Service Provisioning and User Requirements », in *2012 11th International Symposium on Distributed Computing and Applications to Business, Engineering Science*, 2012, p. 191-195.
[30] « Can big data analytics cure the NHS? », *Capgemini UK*, 09-juin-2017. .
[31] « NHS is moving to the cloud with Microsoft », *Microsoft Enterprise - English (en-us)*, 28-juill-2017. [En ligne]. Disponible sur: https://enterprise.microsoft.com/en-us/customer-story/industries/health/nhs-is-moving-to-the-cloud-with-microsoft/. [Consulté le: 21-déc-2017].
[32] Connecting for Health, « Urgent Care Clinical Dashboard Solution Architecture Guide », *Health Developer Network*. .

Prediction and Analysis of Sentiments on Twitter Data using Machine Learning Approach

Ch Srinivasa Rao
Research Scholar, Dept of CSE
Acharya Nagarjuna University, Guntur
Associate Professor, Dept of CS,
SVKP & Dr K S RAJU A&Sc College,
Penugonda, A.P, India
chiraparapu@gmail.com

Dr. G Satyanarayana Prasad
Professor, Dept of CSE,
Dean, Training & Placements
RVR & JC College of Engineering
Chowdavaram, Guntur, A.P, India
Satyam_gp@yahoo.com

Dr.Vedula Venkateswara Rao
Professor, Dept of CSE,
Sri Vasavi Engineering College
Pedatadepalli, Tadepalligudem,
A.P, India
Venkatvedula2017@gmail.com

Abstract—today, methods for automatic opinion mining on online data are becoming increasingly relevant. Over the past few years, methods have been developed that can successfully and with a great degree of accuracy analyze the sentiment in opinions from digital text. These developments enable research into prediction of sentiment. Sentiment prediction has traditionally been used as a tool for stock prediction. In such scenarios, incoming news is analyzed in real-time and the impact of that news on stock prices is estimated, making automatic stock trading possible. Recent developments in sentiment prediction have seen attempts to predict explicit sentiment of the reactions to blogs, before the blogs are even posted. In this paper, we research the prediction of the general sentiment polarity in reactions to news articles, before the news articles are posted. We use Machine Learning approach to solve the sentiment prediction problem. To automatically label comments from Data Set for sentiment prediction training, we perform automatic domain-knowledge transfer from a classifier trained on Twitter data. In this paper, we propose a new machine learning method, a new feature selection method for text and a new machine learning evaluation metric. We provide a thorough analysis of News data, and manually annotate a high standard from it. Finally, we demonstrate the feasibility of sentiment prediction of the general sentiment polarity in reactions to news articles, before the news articles are posted in limited cases. Ultimately, we provide an analysis of the limitations of our and similar approaches to sentiment prediction, and make recommendations for future research.

Keywords- micro blog, twitter, sentiment analysis, opinion mining, predictions

I. INTRODUCTION

On a variety of online platforms, such as review sites, blogs, as well as social services such as Twitter, internet users produce vast amounts of opinionated text about a large range of domains, such as movie reviews, travel experiences, product reviews, opinions about news and others. Automatic opinion mining - the ability to process large amounts of opinionated textual information from online sources without human interference - is necessary. The data sources include opinions about products, brands and developments which increasingly drive the decision making in business and government. Automatic opinion mining is divided into two categories; *qualitative* opinion mining, which attempts to extract pieces of literal information from the data, such as sentences describing an experience relevant to the target of the opinion and *quantitative* opinion mining, which attempts to determine quantifiable dimensions of opinion, such as sentiment. Sentiment analysis is utilized in order to determine the polarity of opinions (positive/neutral/negative) or the emotional charge of opinions across a range of possible emotions (love, fear, anger, understanding etc). The field of sentiment analysis has recently witnessed a large amount of interest from the scientific community [1] [2] [3]. Sentiment analysis has traditionally been applied to a single domain at

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

a time, such as movie reviews or product reviews [4]. More recently, much effort has been invested into development of sentiment analysis methods that can be used across multiple domains [3]. While the creation of general-purpose (cross-domain) sentiment analysis systems remains an unsolved problem, previous advances in sentiment analysis already yield some domain-specific systems which have near-human performance [1] [2]. In addition to sentiment analysis, research into the prediction of sentiment was conducted by a number of researchers [5] [6] [7] [8]. To expand upon the difference between sentiment prediction and sentiment analysis, we consider in abstract detail the methods used in sentiment analysis. Sentiment analysis has been approached from a number of different directions, such as the application of lexicons with manually or semi-automatically annotated word polarities [9], Natural Language Processing methods [10] [11] and machine learning-based approaches [4] [1]. All such approaches determine words or phrases which denote subjective opinion in order to determine the sentiment polarity of the message. For example, [9] uses subjective words annotated with a weight and a polarity, such as "excellent" with a positive polarity and the weight of 5, or "poor", a word with a negative polarity and the weight of 5. Additionally, combined with Naive Bayes machine learning methods as in [4], every word is allocated a particular polarity and weight based on some training examples. Training of such systems is supervised by explicit polarity labeling or implicit interpretation of features such as "smileys" (":)" or "> - (''). In sentiment analysis, the subjective opinion of the creator of a message is explicitly present within the same message. In [11] [10] [12], researchers have established that subjective words, when analyzed using Natural Language Processing methods, are often adjectives, denoting the opinion of the message creator about the noun they belong to. For example "great (ADJ) suspense (NN)" or "worthless (ADJ) plot (NN)", indicate that the object of an opinion is, in many domains, located near the opinion in the message text. On the other hand, when we intuitively consider a domain like news and its character, we observe that news is usually intended to be objective. This means that the opinion of the audience to the themes in a news article is not contained in the article itself. Instead, separate reactions to news, when available, contain the opinions of the audience towards the content of the news. Unlike in sentiment analysis, the text of a news article is not useful to determine the opinions to that article, except perhaps the opinions of the article's author. Since the news article contains objective information, and the opinions to an article are found in the comments to it, the characteristics of the commentators become important. The distinction between sentiment analysis and sentiment prediction is graphically demonstrated in figure 1.1.

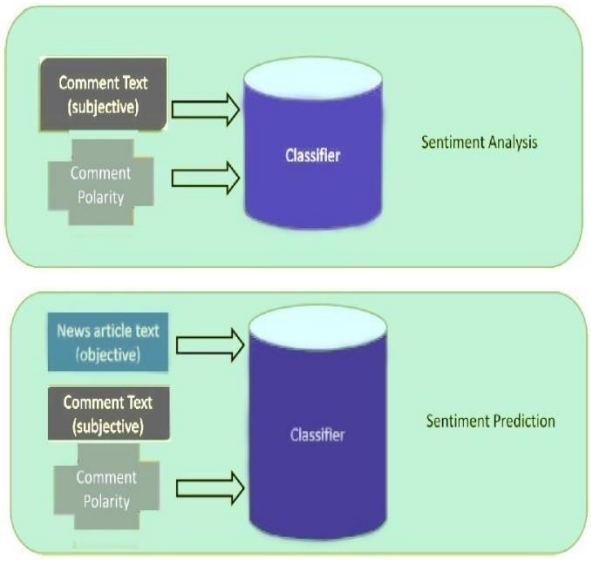


Figure 1.1 the difference between sentiment analysis and sentiment prediction

Automatic prediction of sentiment in reaction to news has already shown merit in practical applications in specific domains. The research focus on sentiment prediction in news, forums and blogs has traditionally been on the prediction of stock movements [5] [6] [13]. Blogs content and reactions are generally automatically processed to produce a reliable stock trading signal: buy, sell or hold. Other domains of sentiment prediction and analysis to news are the generation of news alerts, capturing the general tone in news articles towards relevant issues and assuming this to have predictive properties. Such systems are used to warn its users of "bad news" [14] [15]. In addition, systems have been created that provide the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

context to news articles in terms of their political and emotional polarity [16]. Political sentiment in news has been further explored in the estimation of the emotional impact of political news and blogs to specific audiences [12] [7] [8] [17]. They conclude that, despite the highly personal nature of the motivation behind every individual's reaction to news, in some cases groups of people react to news in similar and predictable ways. This effect is dependent on along which lines people are segmented in groups. The research in this paper focuses on predicting the general sentiment polarity of the reactions to news before a news article is published. News represents a wide domain in which the general sentiment prediction problem is only beginning to be considered [8].

II. Related Work

The research in this paper builds on previous work in the fields of machine learning, feature selection, and sentiment analysis and prediction. In this section we explore a body of work related to these methods.

A. Sentiment analysis

Sentiment analysis can be defined as the automatic extraction of quantitative opinions from subjective text. Over the past number of years, a large body of work describing methods, applications and insights into sentiment analysis was published [1] [2]. Most work has been done on specific domains, such as movie reviews, and on large texts [2], in some cases providing near-human accuracies of sentiment classification. In this section we explore relevant methods for sentiment analysis.

Sentiment analysis is the field of study that analyzes people's sentiment and opinions from written language. It can be carried out at the document level, the message or sentence level or even the aspect or feature level. A popular approach to deal with the job is to pursue a two step approach. In the first step, subjectivity detection, a text is classified as subjective if it represents sentiment, or as purpose if it does not. In the second step, polarity detection, subjective texts are additional classified as positive, negative, neutral or now and then conflict. In a few cases the passion of the sentiment is also considered (Liu, 2012).

According to Taochen et al introduce an interesting approach towards the classification of sentiment in Twitter messages. They described an approach for target extraction and sentence type classification with BiLSTM-CRF. Target extraction is similar to the classic problem of named entity recognition (NER), which views a sentence as a sequence of tokens usually labeled with IOB format. They reduce the feature space in collected tweets by removing the common and uninformative occurrences.

B. Sentiment Prediction

Sentiment prediction can be defined as the automatic prediction of what the quantitative opinions of some audience will be to some message, based on the contents of that message and the earlier observation of a similar audience's response to similar messages. Sentiment prediction is a more difficult task than sentiment analysis, due to the lack of explicit opinion in source data that is classified and the dependency on audience similarity for accurate prediction. Most research into sentiment prediction for news has focused on predicting the movements of the stock market.

C. Sentiment Classification:

Researchers have been working on diverse characteristics of sentiment analysis and projected a range of algorithms and methods. Generally, there are two approaches for sentiment analysis: the lexicon based approach and the machine learning approach, as shown in Figure 2.1.

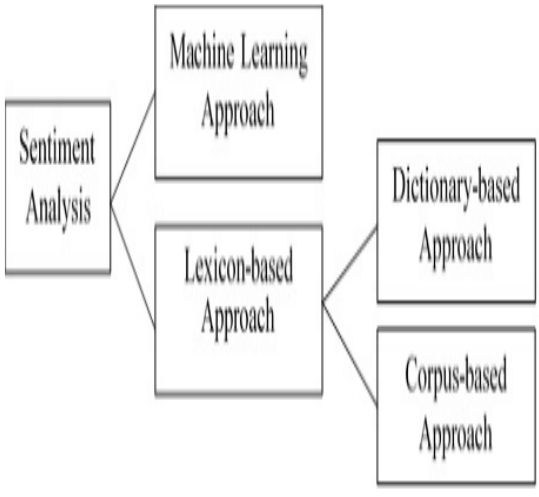


Figure 2.1 Sentiment Classification Techniques

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

D. Lexicon based approach:

For the lexicon-based approach, as the name implies, sentiment lexicon is the main source in the process of sentiment analysis, such as collections of sentiment phrases or dictionaries of sentiment words. Actually, lexicon-based methods are divided into two classes, which are dictionary-based and corpusbased [Serrano et. Al.]. Dictionary-based methods mainly rely on dictionaries which are collected and annotated for sets of common opinion terms. Although it is a reliable solution for sentiment analysis, there are some limitations when dealing with different domains or analyzing contexts. On the other hand, corpus-based methods are also based on dictionaries, but these dictionaries contain related opinion terms or phases for specific fields.

For the task of general sentiment analysis, which determines the sentiment of a given text, such as a movie review or a product review, a variety of lexicon-based methods have been proposed. Turney provided a general algorithm to analyze online reviews simply as recommended (thumbs up) or not recommended (thumbs down). The average semantic orientation had been used as a tool to classify opinions from adjective or adverb phases of reviews [turney et. Al.].

E. Machine Learning approach:

Machine learning is a part of Artificial Intelligence. It includes two main categories: supervised learning and unsupervised learning [Serrano et. Al.]. Machine learning algorithms deal with different tasks to allow computers to learn. Usually, machine learning algorithms work well on inferring information about the properties of sets of data.

Different machine learning techniques might be used to handle different sentiment classification problems. From the previous work, many researchers chose machine learning approaches to deal with their tasks of sentiment analysis. They experimented different machine learning algorithms, such as Naive Bayes classifier, Support Vector Machine (SVM) classifier, Maximum Entropy classifier and so on. Because of using different pre-processing methods and different training data, these classifiers can give different performance. By far, the majority of the research work on sentiment analysis from the machine learning approach was focused on relatively and highly subjective English normal texts, such as product reviews and movie reviews. On the other hand, there has been also some research on sentiment analysis of tweets that are short, messy and multilingualism.

III. Predictive Model and Architecture

In this section we describe our proposed learning method for predicting the popularity of tweets in twitter social networks. We model the problem of popularity prediction as a binary classification problem. Our proposed approach to popularity prediction is based on a feature-based classification model in which we extract a set of features from tweets and classify them as popular/unpopular classes. In this section we introduce the overall architecture of our system, which comprises different components. Figure 3-1 illustrates the overall architecture of our proposed model. The data collection method is illustrated in the left side of the figure. The model then extracts several features from tweets and different machine learning approaches are used to train classifiers. The classifier is then used to predict whether a tweet will be popular or not. The two important decision for earning-based systems are the choice of classifier and the features that are extracted from the data.

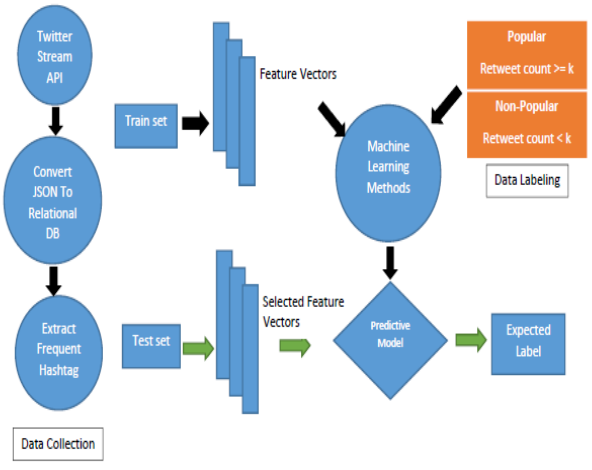


Figure 3.1 Schematic representation of model architecture

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

IV. Proposed Methodology

In this section, we conduct an extensive and systematic empirical study on machine learning algorithms for tweet sentiment analysis. Different from previous research works, we compare the performance of 5 different machine learning algorithms using 4 popular text processing methods based on 12 datasets. For deriving more accurate results, we use 10 times of 5-fold cross validation technique to run our experiments.

A. *General Idea*

For our empirical study, choosing machine learning algorithms is the most important part. However, it is not easy to decide which algorithm we should use, because there are a lot of machine learning algorithms that have been introduced by numerous researchers. Algorithm selection depends on many various conditions, such as size and type of data, hardware, time cost, etc. Even most researchers cannot tell which algorithm is the most suitable before trying them. As the motivation of this study is to investigate the performance of machine learning algorithms on tweet sentiment analysis, we select widely used machine learning algorithms for a systematic comparison: Naive Bayes (NB), Support Vector Machines (SVMs). We believe that these five algorithms are highly representative.

B. Machine Learning Algorithms

B.1 Naive Bayes

Naive Bayes is an algorithm which is widely used for classification problems. In classification issues, a classifier is built from a training data set and it will classify data into a number of different classes. The training data set consists of a number of labeled training examples. Each training example E is represented by an attribute vector (a_1, a_2, …. a_n) and the class c for the vector. For each attribute a_i in the vector, it represents the value of attribute A_i. For example, when we want to describe the weather for today, we can set a number of weather attributes, such as: outlook, temperature, humidity. The class label c of each training example is given, where c is the value of the class variable C. The built classifier will be used to predict the labels for any new instance. Naive Bayes is based on the so-called conditional independence assumption. The attributes are independent between each other given the class label, shown in the equation below.

$$P(E \mid c) = \mathrm{p}(a1, a2\ldots \mid c) = \prod_{i=1}^{n} p(ai \mid \mathrm{C})$$

Where $p(a_i \mid c)$ stands for the likelihood of Ai, and instance $E = (a1, a2, \ldots, an)$. In the task of classification, for each instance E, the value of $p(E)$ is a constant. So the probability $p(c \mid E)$ is:

$$p(c|E) \propto p(c) \prod_{i=1}^{n} p(a_{i|c})$$

Given an instance E, the naive Bayes classifier is defined as below:

$$NB(E) = \underset{C}{argmax} P(c) \prod_{i=1}^{n} P(a_i|c)$$

Accordingly, learning naive Bayes involves mainly estimating the values of the probability $p(a_i \mid c)$ from the training instances. That is, the values of the probability $p(a_i \mid c)$ will be computed using the training data, for each value a_i of the attribute Ai given the value c of the class variable C.

Naive Bayes has been well investigated. The major benefits for Naïve Bayes are listed below.

- It is not complicated and speedy to predict the class of any new instance.
- It also performs superb in multi-class prediction.
- When the pre-assumption of conditional independence holds, Naïve Bayes classifier performs better comparing to other models, such as logistic regression, using less training data.

On the other hand, there are also some technical issues in applying Naïve Bayes. The major limitation of Naive Bayes is the conditional independence assumption.

- In real life, it is almost impossible.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

- If there is a new attribute value or class label for some instances in a testing data set not observed in the training data, the model will assign a 0 (zero) probability and this is non-advantageous to make a good prediction. This is often known as “Zero Frequency”. In order to solve this issue, we have an option to use the smoothing technique. One of the simplest smoothing techniques is called Laplace estimation.
- Naive Bayes is also known as an unremarkable estimator. The probability outputs from predict probability are not accurate.

B.2 Support Vector Machine

Support Vector Machines (SVMs) are happening to be exceptionally accomplished as text categorization, widely outperforming Naïve Bayes. SVMs uses a function called kernel which are machine learning classification methodology in which data is not separable linearly in the new area which it is to locate to area of data points, with allocation for classification of erroneous.
Support Vector Machines are the members of the family of classifiers which are linear. The main objective of linear classifier is to find hyper plane which is linear in nature of feature area that divides all other entries in the form of two classes. The main function of SVMs is to find out the hyper plane which is separating that has distance maximum from the nearest points to feature area in it. Searching hyper plane in sample of linear separable, the equation can be considered as problem of optimization.

$$\| \omega \|^2 \rightarrow \min (\omega, b)$$

$$y_i (W^T X_i + b) \geq 1, j=1 \ldots m$$

Here is the area between the points of both second and first class and the hyper plane and it is nearest to $y_i (W^T X_i + b)$, the product of its position relative to the hyper plane and point class value.

V. Experimental Resuls and Analysis

A. Implementation

We are extracting tweets from the twitter with the help of the Java API called Twitter4j. It consists various number of libraries that are used in the extraction. At first we have added this library into our java project. Then with the help of twitter app we have obtained Consumer Token Key and Access Token Key. Further, extraction of tweets will be start only after when we generate Access Key.

Generation of Access Key needed every time for the extraction of the tweets. The twitter4j containing libraries are shown in Figure 5.1.

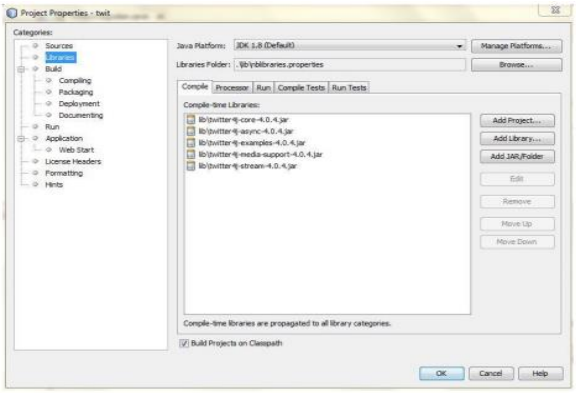


Figure 5.1 Twitter4j libraries

Consumer Token Key will be provided by the twitter app. There is a unique key for every app and that key known as Consumer Token Key. In order to obtained tweets we have to apply consumer token key and access token key into the java code.

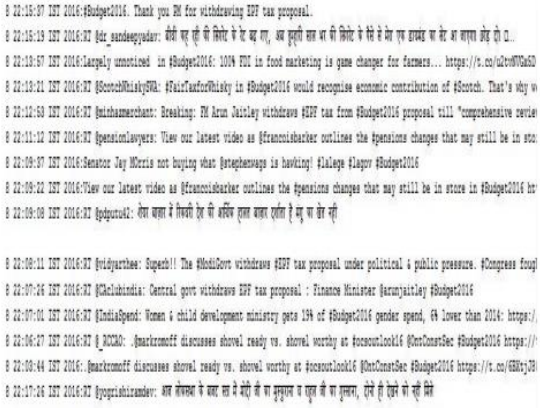


Figure 5.2 List of Tweets

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

B. Preprocessing using R

In this sectioncollected data is pre processed. We have used R language for the pre processing. Stop words, user references, urls etc are removed from the data. Regular expressions are used to remove url. Collected tweets are then manually labeled and stored in files as test dataset. We have two data sets: positive and negative. We have created two separate files for positive and negative set as shown in Figure 5.3 and Figure 5.4.

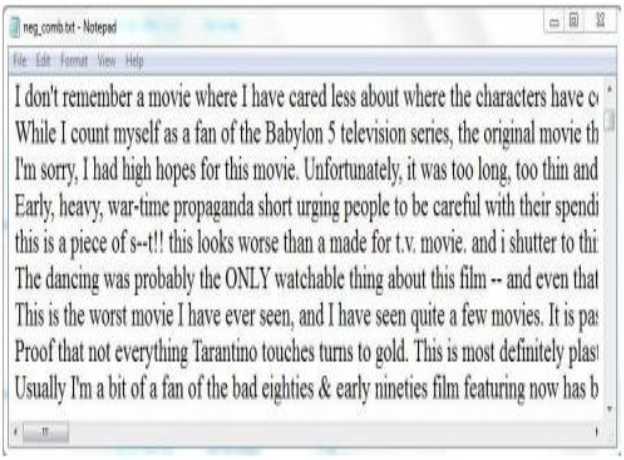


Figure 5.3 Positive Training Set

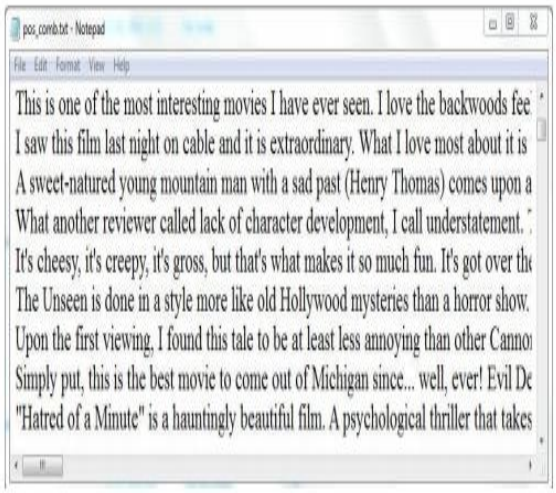


Figure 5.4 Negative Training Set

C. Results

We are using R language for implementation. R language offers maximum support when it comes machine learning techniques. Machine learning techniques can be easily implemented in R language. Packages that we are using are "RTextTool","Rweka"and "e1071". RTextTools have most of the machine learning algorithms but not have Naïve Bayes, which is included in e1071 package and Rweka package is used for n-gram feature.

The reason we are using R language because when the dataset is big, it is fast and efficient in terms of performing. The packages in the R tool are updated regularly and have greater number of probabilistic and statistical functions.

D. Result Analysis

To calculate the accuracy of classifier we required measure on which accuracy can be obtained. There are two measures on which accuracy can be dependent:

- Precision
- Recall
- Accuracy

Let's take collection of M documents, M_P denotes the number of document which belongs to the true positive class and M_N denotes the number of documents which belongs to the true negative class. TP documents had rightly classified whereas FP documents are wrongly classified, similarly FN documents are wrongly classified and TN documents are rightly classified.

Precision: It is the ratio of documents of rightly classified under positive prediction class to all documents under positive prediction class.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: It is the ratio of documents of rightly classified under positive prediction class to the documents that are positive in the negative prediction class.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Recall = $\frac{TP}{TP+FN}$

The following Figure shows conclusion Matrix

		True Class: Positive	True Class: Negative
Prediction Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 5.5: Confusion Matrix

Accuracy: In order to check which n-gram feature will give better results for these three models, we have to find the accuracy of classifiers. Accuracy for any prediction model can be given as

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

The following tables show the Precision and recall for different features like unigram, bigram and hybrid

Table 5.1: Precision and recall for Unigram feature

Algorithm	Unigram	
	Precision	Recall
Naïve Bayes	0.75	0.71
Support Vector Machines	0.82	0.76

Table 5.2: Precision and recall for Bigram feature

Algorithm	Bigram	
	Precision	Recall
Naïve Bayes	0.72	0.70
Support Vector Machines	0.76	0.71

Table 5.3: Precision and recall for Hybrid feature

Algorithm	Hybrid	
	Precision	Recall
Naïve Bayes	0.73	0.71
Support Vector Machines	0.83	0.74

We have obtained the result as hybrid feature with svm classifier gives the best results for pereddiction of sentiment of twitter data. We obtained 84% accuracy using hybrid feature on svm classified data. 70% is least we have obtained in bigram with Naïve Bayes classifier. The results can be shown in Figure 5.6

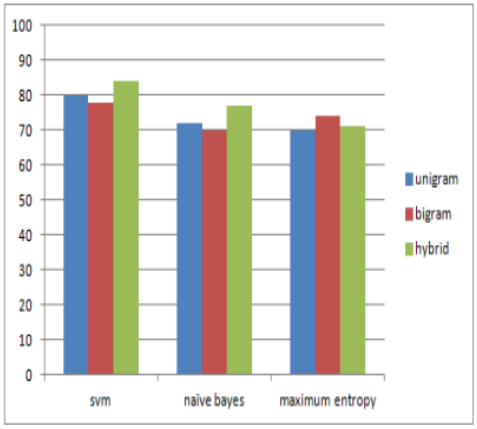


Figure 5.6 Results of machine learning algorithms

VI. Conclusion

In this paper, we have implemented supervised classifiers like Naïve Bayes, support vector machines and maximum entropy using unigram, bigram and hybrid (unigram + bigram) feature. There is need to do sentiment analysis as texts in form of messages or posts to find the whether the sentiment is negative, positive or neutral. We had extracted data from twitter i.e. movie reviews for sentiment prediction using machine-learning algorithms. First we extracted the data from twitter using twitter API. Then in pre-processing, we clean the data and make the data available to train using classifiers. We have collected 15000 tweets for training set and 2000 tweets for testing set. SVM using hybrid feature outperforms all other classifiers and selection feature with accuracy of 84%. It is

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

concluded that SVM gives better results than other classifiers.

In future, we are planning to make automatic sentiment classifier for more than one languages starting from the Hindi language. As nowadays multilingual messages are posted in twitter, so we will able to predict the sentiment for any language.

REFERENCES

J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, Sentiment analysis: a review and comparative analysis of web services, Information Sciences, 2015, Vol.311, pp.18–38.

Walaa Medhat , Ahmed Hassan , Hoda Korashy , Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal (2014) 5, 1093–1113

Marouane Birjalia, Abderrahim Beni-Hssanea , Mohammed Errital, Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks, ScienceDirect Available online at www.sciencedirect.com Procedia Computer Science 113 (2017) 65–72

Taochen, Rruifengxu, Yyulanhe, Xxuanwang, Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN, Expert Systems with Applications, Volume 72, 15 April 2017, Pages 221-230

Ali Hasan , Sana Moin , Ahmad Karim and Shahaboddin Shamshirband, Machine Learning-Based Sentiment Analysis for Twitter Accounts, Math. Comput. Appl. 2018, 23, 11; doi:10.3390/mca23010011

Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. Ain Shams Eng. J. 2014, 5, 1093–1113. [CrossRef]

Sebastiani, F. Machine learning in automated text categorization. ACM Comput. Surv. 2002, 34, 1–47. [CrossRef]

Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M. Lexicon-based methods for sentiment analysis. Comput. Linguist. 2011, 37, 267–307. [CrossRef]

Prabowo, R.; Thelwall, M. Sentiment analysis: A combined approach. J. Informetr. 2009, 3, 143–157. [CrossRef]

Dang, Y.; Zhang, Y.; Chen, H. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. IEEE Intell. Syst. 2010, 25, 46–53. [CrossRef]

Cambria, E. Affective computing and sentiment analysis. IEEE Intell. Syst. 2016, 31, 102–107. [CrossRef]

Jagdale, O.; Harmalkar, V.; Chavan, S.; Sharma, N. Twitter mining using R. Int. J. Eng. Res. Adv. Tech. 2017, 3, 252–256.

Anjaria, M.; Guddeti, R.M.R. Influence factor based opinion mining of twitter data using supervised learning. In Proceedings of the 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), Bangalore, India, 6–10 January 2014; pp. 1–8.

Dubey, G.; Chawla, S.; Kaur, K. Social media opinion analysis for indian political diplomats. In Proceedings of the 2017 7th International Conference on Cloud Computing, Data Science & Engineering, Noida, India, 12–13 January 2017; pp. 681–686.

Liu, B.; Hu, M.; Cheng, J. Opinion observer: Analyzing and comparing opinions on the web. In Proceedings of the 14th International Conference on World Wide Web, Chiba, Japan, 10–14 May 2005; pp. 342–351.

Razzaq, M.A.; Qamar, A.M.; Bilal, H.S.M. Prediction and analysis of pakistan election 2013 based on sentiment analysis. In Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, China, 17–20 August 2014; pp. 700–703.

Soucy, P.; Mineau, G.W. A simple knn algorithm for text categorization. In Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 29 November–2 December 2001; pp. 647–648.

Lewis, D.D. Naive (bayes) at forty: The independence assumption in information retrieval. In Proceedings of the 10th European Conference on Machine Learning, Chemnitz, Germany, 21–23 April 1998; pp. 4–15.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Segnini, A.; Motchoffo, J.J.T. Random Forests and Text Mining. Available online: http://www.academia. edu/11059601/Random_Forest_and_Text_Mining (accessed on 26 February 2018).

Raschka, S. Naive bayes and text classification i-introduction and theory. arXiv 20

H. Kang, SJ. Yoo and D. Han, " Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews", Expert Systems with Applications, Elsevier, vol. 39, no. 5, pp. 6000—6010, 2012.

Ch Srinivasa Rao is a Research Scholar in the Department of Computer Science & Engineering at Acharya Nagarnuna University, Guntur, A.P, India. He is working as Associate Professor in SVKP & Dr K S Raju A&Sc College, Penugonda, A.P. He received Masters Degree in Computer Applications from Andhra University and Computer Science Engineering from Jawaharlal Nehru Technological University, Kakinada, India. His research interests include Data Mining, Big Data Analytics.

Dr.G Satyanarayana Prasad is Professor in the Department of Computer Science Engineering and Dean, Training & Placements at RVR & JC College of Engineering, Chowdavaram, Guntur, India. He received M.S in Computer Science from A&M University, ALABAMA, USA and PhD from Andhra University, Visakhapatnam, India. His research interests include Image Processing, Data Mining, Big Data Analytics. He guided two research scholars for the award of their PhD. He published books, several papers in International conferences and journals.

Dr.Vedula Venkateswara Rao is Professor in the Department of Computer Science Engineering at Srivasavi Engineering College, tadepalligudem, India. He received Matsers Degree in Computer Science Engineering from JawaharLal Nehru Technological University Kakinada, Masters Degree In Information Technology from Punjabi University, Patiayala, India and PhD from Gitam University. His research interests include Cloud Computing and Distributed Systems, Data Mining, Big Data Analytics and Image Processing. He published several papers in International conferences and journals.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Enhancement in Noise Removal Techniques by Using Hybrid MediangausTransform Method for Paddy Seeds

Dr.(Mrs).M.Renuga Devi,
Director, Department of Computer Applications,
Sri Venkadeswara College of Computer Application and Management, Coimbatore.

Mrs.S.Maheswari, Ph.D Scholar,
Bharathiar University, Coimbatore.

Abstract:

The aim of this research is to develop a technique for identifying different variety of paddy seeds by using morphological features these morphological features are very effective in identifying the different variety of paddy seeds. The image of the paddy seeds are acquired by the cannon digital camera for high quality images. These captured images are stored in the jpg format for future process. The research involves image acquisition, segmentation, feature extraction and classification. This paper focuses on pre-processing by removing the noise which accorded during image acquisition by applying hybrid mediangaustransform method.

Keywords: pre processing, RGB model, seed Region separation.

Introduction:

Paddy plantation is one of the most important crop cultivation in the agricultural countries. The plantation involves lot of aspects in due consideration for good yield. The most important among those is the type of seed used for cropping. Others attributes involve water fertile soil fertilizers etc., paddy is one of the cereal crop and staple food for most of the people in the world. In India paddy is the most important cereal gain crop[1]. For tropical Asians it is the staple food and is the major source of dietary energy and protein. In Southeast Asia alone, paddy is the staple food for 80% of the population. During the cultivation information about the gain type, grain quality is needed at each stage in hand before the start of the next process. So that the next course of work can be determined and performed. Measurement of some characters such as color, texture, or some of morphological features are simple, but the information which get in this way is subjective and so is not reliable. Therefore, digital image analysis is the method to solve the problem. This method will give accurate, fast and much reliable information [3].

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Proposed Methodology:

Desirable part of research related to Image processing is Image Pre-Processing. Pre Processing of image enhances the quality of image by underlying various steps which would likely to remove the image distortions and to enrich the corresponding features of the image to produce better results.

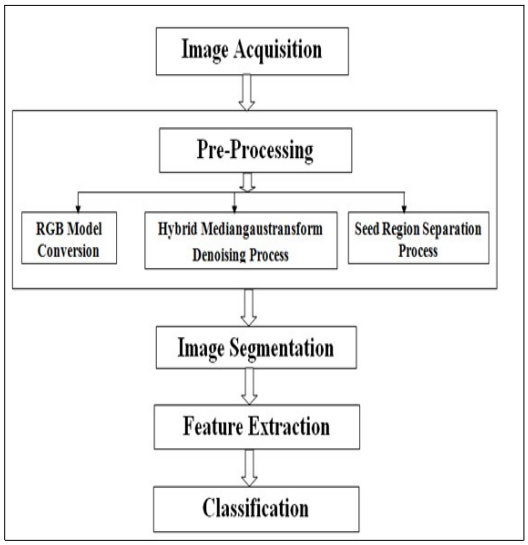


Fig:1:Procedural Flow

In the resultant paper, technique is used to Seed image database of 3 uint8) into 256 x 256 process the image is can be coupled with common color map. 'bicubic' interpolation resize the original RGB pixels size (1109 x 1069 x dimensions. Following the identical in size moreover it indexed images on a

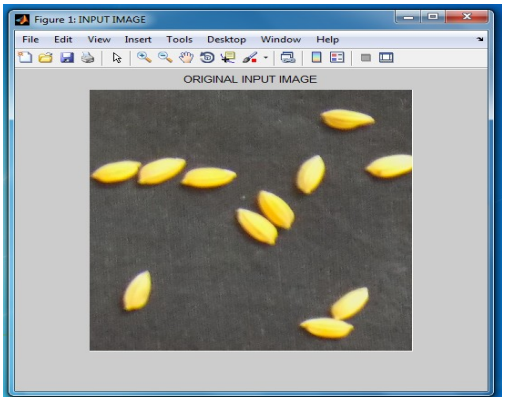


Fig: 2: Original Input Image

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Further the pre processing method is divided into three stages namely, (i) RGB model Conversion (ii) Image denoising (iii) Seed region separation process. It is described in the figure: 1. Stage one involves the process of Color model conversion of RGB to Ycbcr model is extracted. Later, the Ycbcr image having 3 dimensional parts (i.e., 'Y' is essentially a greyscale or luma copy of the original image, cr is chrominance red, cb is chrominance blue). The 'Y' model or grayscale portion is taken into further process.[14] In this gray portion is to convert into standard double precision value (figure: 3). Stage two introduces the ***hybrid mediangaus transform method.***

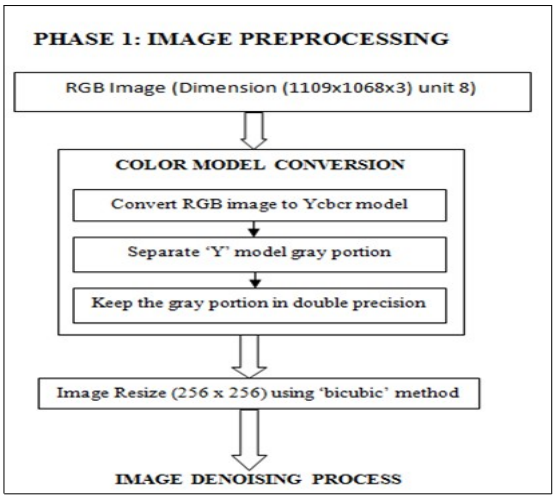


Fig:3:Image Pre Processing–colorModel Convertion

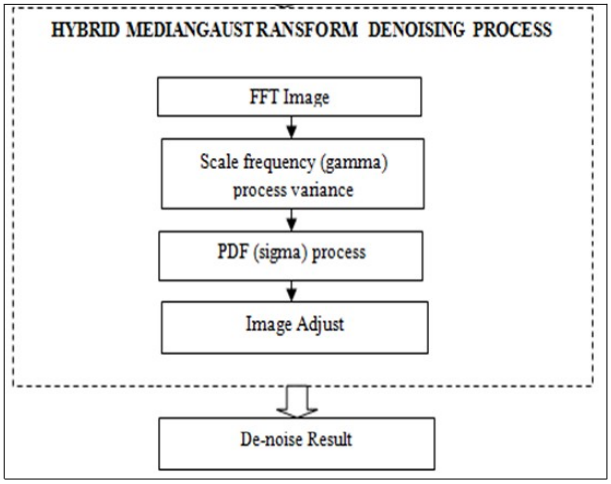


Fig:4: Hybrid Mediangaustransform Denoising

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The proposed ***hybrid mediangaustransform*** method (figure: 4) produces 0.5 noise variation (sigma) value, luminance correction (gamma) value is 0.5 and it also calculates the height of the image. Further the image is converted using Fast Fourier Transform method hence it sets its initial parameter to FFT.[10] The resulted image is converted to Inverse FFT Image with pixel variation by calculating Scale (gamma), High Frequency for Image Variations and Probability Density Function for sigma. The acquired de-noised image is displayed using image adjust function (fig: 5). The proposed ***hybrid mediangaustransform*** shows highest PSNR value compared to median filter, Gaussian filter and wiener filter.

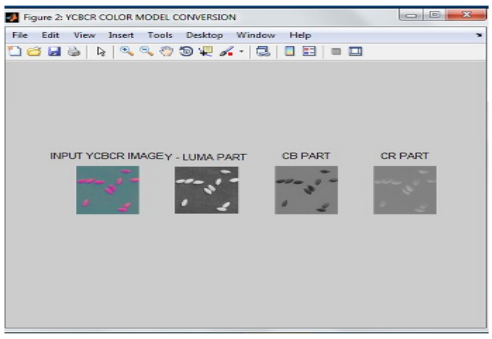


Fig:5: Output of Color model Conversion

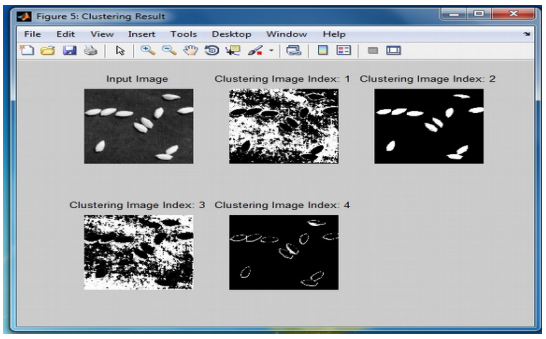


Fig:6: Clustering

Image acquired from stage two is passed into next stage for seed region separation. In this stage the seed region is separated from the de-noised image. It includes three steps a) image orientation, b) clustering and c) seed object properties.

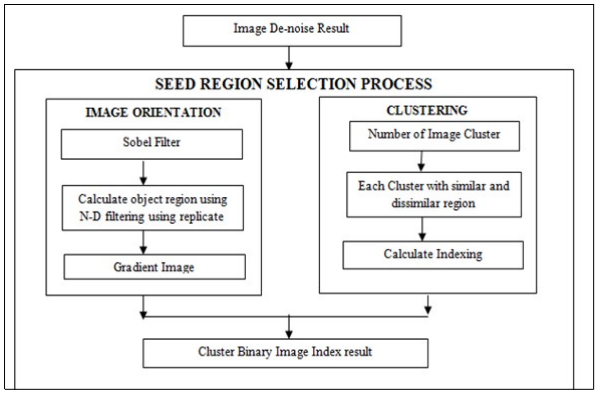


Fig: 7: Detailed flow of Seed Region selection.

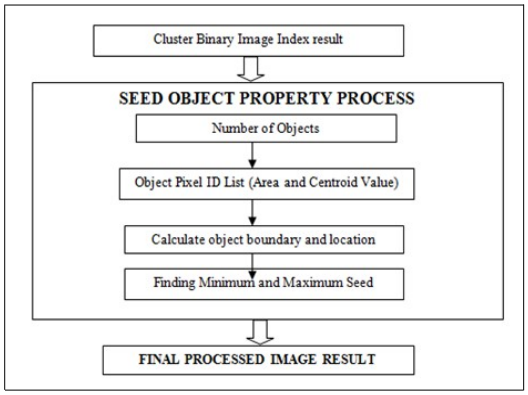


Fig:8:Calculation of Clustering Index-Flow

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

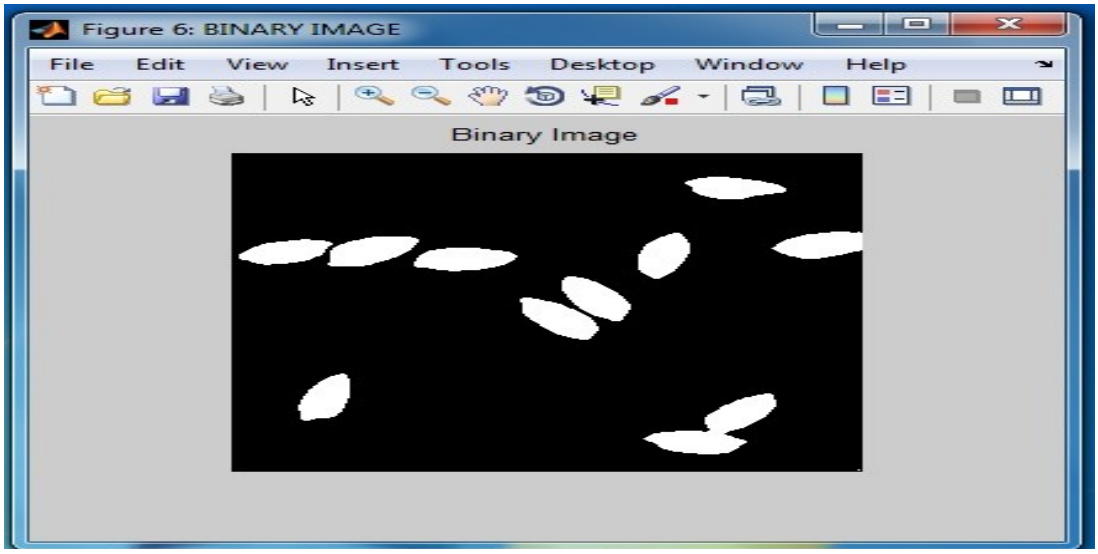


Fig:9: Binary Image

In first step image orientation is done using sobel filter to produce a gradient image. Image cluster is counted by clustering the similar and dissimilar region to find the cluster index. It produced cluster index binary image (figure:9). The obtained image is used to calculate the number of objects, area, centroid value, object boundary and location, minimum and maximum size seed and individual seed and displayed it (Table: 1 and figure: 10).

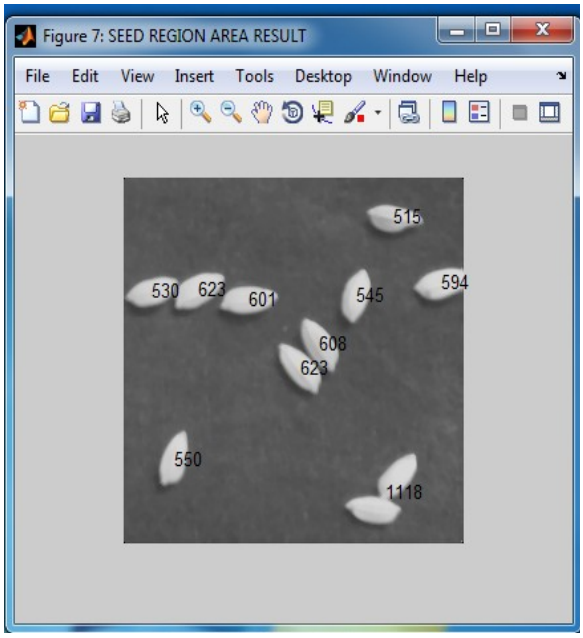


Fig:10: Output of seed region selection

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

ImageSize: [256 256] **NumObjects: 10** **PixelIdxList: {1x10 cell}**		
No	**Area**	**Centroid**
1	530	[22.5 79.88]
2	550	[38.78 197.65]
3	623	[56.91 78.98]
4	601	[94.68 85.69]
5	623	[133.56 133.62]
6	608	[148.28 117.36]
7	545	[175.82 82.81]
8	1118	[197.9 221.32]
9	515	[203.84 28.4]
10	594	[239.84 74.4]

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table: 1: Clustering Index values

The following is the proposed Hybrid Mediangaustransform denoising algorithm.

Input: Image ***I***, gamma $\gamma = 0.5$, sigma $\sigma = 0.5$, Scale Frequency ***sf***, height ***h***, Gaussian ***Gf.***

Output: Denoise Image ***DR***.

Initialization

Step 1: Calculate image dimensions

Step 2: Convert Image using FFT Conversion

Step 3: Define Fast Fourier Transform for Initial Parameters (height and image total number pixels (***N***))

Step 4: Calculate Absolute probability density function (pdf) value of FFT Image with Total Number of Pixels

Process

Step 5: Calculate image dimension noise scale frequency

$$sf = fft(Total\ pixels) \times \left\| (fft(N) > 0) \right\| + \frac{1}{\square} \times \left\| (fft(N) = ¿0) \right\| eqn.(1)$$

Step 6: Calculate Probability Density Function for noise variation sigma

$$pdf = pdf \times (pdf > \sigma^2) + \sigma^2 \times (pdf \le \sigma^2) eqn.(2)$$

$$Gf = \frac{sf \times (pdf - \sigma^2)}{pdf - (1-1) \times \sigma^2} eqn.(3)$$

$$Result = Gf \times fftimage\ eqn.(4)$$

Step 7: Denoise Result

$$DR = real(ifft(Result))\ eqn.(5)$$

Result and discussion:

The process of de-noising the image is split into three stages as RGB model Conversion, Image denoising and Seed region separation process. In each stage it is discussed above and their results were disclosed. Table1 shows the cluster index and the seed region selection is studied in

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

figure 10. The result showed variance in the noise present the image. The proposed ***hybrid mediangaustransform*** shows highest PSNR value compared to median filter, Gaussian filter and wiener filter.

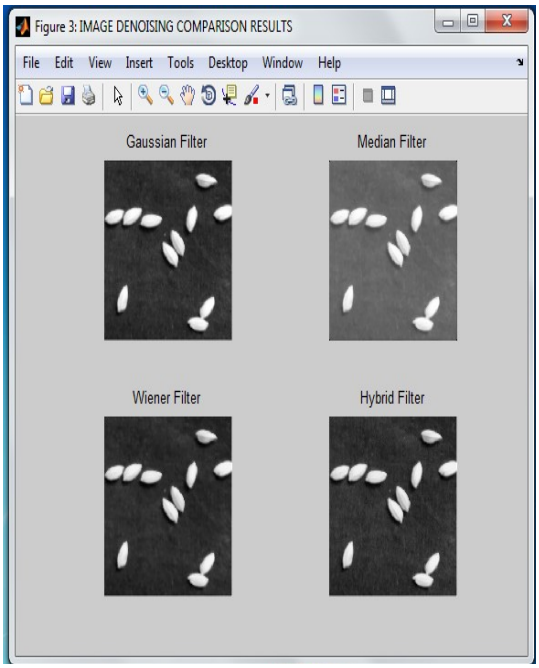


Fig:11: Output for different filters.

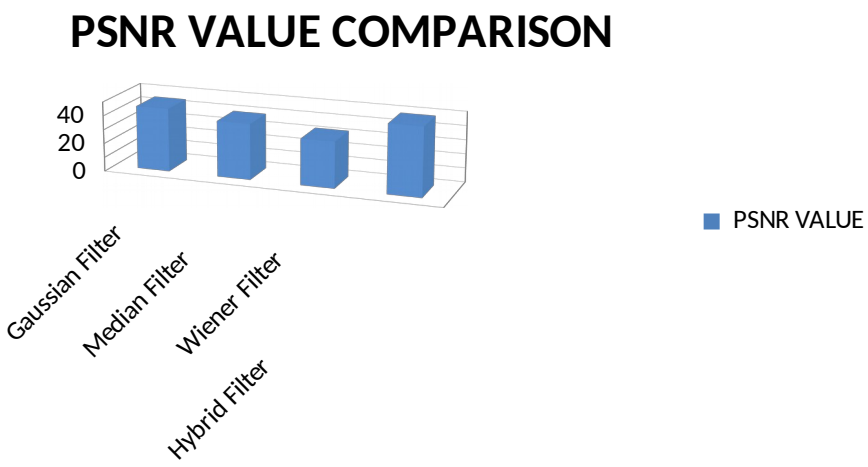


Chart:1: PSNR Value Comparison

DENOISE FILTERING METHODS	PSNR VALUE
Gaussian Filter	45.0419
Median Filter	39.6421

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Wiener Filter	32.8538
Hybrid Filter	48.9129

Table: 2: PSNR Value

The chart: 1 provides the evidence; result of the posed algorithm shows the better results when compared with other filters like Gaussian Filter, Median Filter Weiner Filter(figure:11).Hybrid mediangaustransform algorithm show high PSNR value.

Conclusion:

The original image is converted into RGB colour model conversion for the further process of the image. The obtained image is passed into the proposed method were it produced produces 0.5 noise variation (sigma) value, luminance correction (gamma) value is 0.5 and it also calculates the height of the image The acquired de-noised image is displayed using image adjust function. The proposed ***hybrid mediangaustransform*** shows highest PSNR value compared to median filter, Gaussian filter and wiener filter.

References:

[1]. Rubi Kambo, Amit Yerpude, "Classification of Basmati Rice Grain Variety using Image Processing and Principal Component Analysis", May 2014.

[2]. Davinder Sandhu, "Image Segmentation based Methodology for Classification of various Seed varieties.", April-May 2013.

[3]. Sandeep Arya and Parveen Lehana, "Development of a Seed Analyzer using the Techniques of Computer Vision", January 2012.

[4]. Archana Chaugule and Suresh N. Mali, "Evaluation of Texture and Shape Features for Classification of Four Paddy Varieties" 18th August 2014.

[5]. Leila Farahani, "Discrimination of some cultivars of durum wheat (Triticum durum Desf.) using image analysis", 2012.

[6]. M. S. Howarth, P. C. Stanwood "Measurement of Seedling Growth Rate by Machine Vision", may-june-1993.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[7]. Dell, Aquila "Application of a Computer–Aided Image Analysis System to Evaluate Seed Germination under Different Environmental Conditions" Ital. J. Agron., 8, 1, 51-62,2004.

[8]. I. Zayas, Y. Pomeranz, and F. S. Lai, "Discriminate between wheat and non-wheat components in a grain Sample", Cereal Chem., vol. 66, no.3, 1989.

[9]. N. S. Visen, D. S. Jayas, J. Paliwal, and N. D. G.White,"Comparison of two neural network architectures for Classification of singulated cereal grains", Can. BioSyst. Eng, vol. 46, 2004.

[10]. J. Paliwal,M. S. Borhan and D. S. Jayas, "Classification of cereal grains using a flatbed scanner", Can Biosyst Eng,vol. 46, 2004.

[11]. M. A. Shahin and S. J. Symons, "Seed sizing from images of non-singulated grain samples",Can. BioSyst. Eng, vol. 47, 2005.

[12]. P. M. Granitto,H. D. Navone, P. F. Verdes, and H. A. Ceccatto, "Automatic identification of weed seeds by color image processing",2000.

[13]. H. Rautio and O. Silvn, "Average Grain Size Determination using Mathematical Morphology and Texture Analysis",2000.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Deep Learning based Vehicle Detection and Tracking Techniques: State-of-the-Art Survey

Vikram Kumar
Research Scholar, Department of Computer Science
Thapar Institute of Engineering and Technology, Patiala, Punjab, India
*Email:*vikramthakur2623@gmail.com

Ashima Singh
Assistant Professor, Department of Computer Science
Thapar Institute of Engineering and Technology, Patiala, Punjab, India.
Email: ashima@thapar.edu

***Abstract*——Vehicle detection has become anessential task because of the rising usage of surveillance cameras in smart cities, road network managent, highway and urban traffic planning etc.But detection of vehicles faces many challenges such as occluded vehicles, shadows of structures, similarity in designs of vehicle leading to classification issues. Deep Learning based algorithms such as CNN, RCNN, Faster CNN etc. provides appropriate solution to facilitate vehicle detection because of the self learning capability of the algorithm after training. This paper aims to present an overview of various vehicle detection techniques based on deep learning which can effectively be used for video surveillance in highway, road and traffic management systems.**

Keywords—CNN, RCNN, Faster CNN, Deep Learning, Vehicle Detection, Occlusion.

I. INTRODUCTION

Vision-Based Intelligent Transportation System (ITS) is one of the critical applications of video-based supervision frameworks. To disentanglevaluable and exact vehicle movement data, some of the most researched topics are transportation management and their activity building applications. The data for examination of activity in picture and movement stream control data like vehicle direction, vehicle following, vehicle stream, vehicle order, movement thickness, vehicle speed, movement path changes, and so forth. Previously, automation of toll-collect framework used vehicle recognition, division to decide the charge for different kind of vehicles. As of late, vehicle acknowledgment framework is utilized for: 1 identification of the vehicles; 2 detection of the activity paths; 3 ordering the kind of vehicle class on expressways and streets. The customary vehicle frameworks witnessed great decrease and not perceived well as the vehicles usually are blocked by different vehicles or by foundation impediments like street signals, trees, and climate conditions. Moreover, the executions of these frameworks rely upon a decent activity-picture examination ways to deal with recognization, tracking and characterization of the vehicles. This review paper explores the various deep learning based vehicle detection techniques used in applications such as traffic surveillance, battleground survey, autonomous vehicles and aerial survey.

II. LITERATURE SURVEY

Tayara et al. [1] proposed vehicle detection and tallying framework for airborne applications.The configuration uses convolution neural networks to relapse a vehicle spatial thickness outline the aeronautical picture. A completely convolutional relapse network is utilized as a part of this framework whose objective is to limit the error between the ground truth and anticipated yield. It comprises of two ways: up-inspecting way and down-testing way. The down-examining way is pre-prepared VGG-16 network which has rehashed cushioned 3x3 convolutionsfollowed by a maximum pooling activity. They use upto layer conv5 from VGG-16 networkandrest of layers are removed to reduce the number of parameters. The up-sampling path symmetry is removed due to the assymetric nature of image regression. A 2D circular Gaussian function is used for creating the ground-truth from the dataset. So as to expand the quantity of preparing illustrations, information expansion methods, for example, revolution, horizontall and vertical flipping and moving are utilized. At last, a basic associated segment calculation is utilized to find the areas and count of the blobs. The proposed framework is assessed on two open datasets to be specific DLR Munich Vehicle dataset gave by the German Aerospace Center and Overhead Imagery Research Data Set (OIRDS) dataset. The proposed framework can recognize the vehicles in flying pictures precisely with better accuracy scores. The research also devours additional time amid derivation contrasted with different frameworks.

Chu et al.proposeda novel vehicle detection scheme in view of multi-assignment profound Convolutional Neural Networks (CNNs) and Region of Interest (ROI) voting. In the design, the administered data is enhanced with subcategory, jumping box relapse, region cover, and class of each training ROI as a mutli-assignment learning-based framework.This configuration enhances the adequacy of powerful detection by permitting the CNN model to share visual information among various vehicle traits at the same time. This outline uses the CNN model to anticipate the offset course of every rous limit towards the relating ground truth. A subcategory NMS strategy is acquainted with

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

conquered the difficulties confronted while having impeded vehicles. At that point every rous can vote in favor of appropriate contiguous bouncing boxes that are reliable with the extra data. The scores are contrasted and the scores of the individual ROI itself to acquire more exact outcomes. Test comes about are discovered utilizing KITTI PC vision benchmarks and the PASCAL 2007 vehicle dataset is utilized for preparing. The outcomes showssuperior execution to existing distributed works.

Cai et al. [3] suggested a versatile scene vehicle detection calculation in view of composite profound structure. This outline is propelled by the Bagging (Bootstrap Aggregating) instrument. Various generally autonomous source tests are first used to fabricate different classifiers and after that voting is utilized to create target preparing tests with certainty scores. The programmed highlight extraction capacity of profound convolutional neural networks is then used to perform source-target scene include comparability clcultions with a depp auto-encoder keeping in mind the end goal to outline a composite profound structure based scene-versatile classifier and its preparation technique. Tests on the KITTI dataset and a dataset caught by the gathering exhibit that the proposed strategy performs superior to anything existing machine-learning based vehicle detection strategies. What's more, contrasted and existing scene-adpative protest detection techniques, the proposed strategy enhances the detection rate by a normal of roughly 3%.

Deng et al. [4]explored Region-based convolutional neural networks (R-CNNs) have a few difficulties when utilized for vehicle detection in ethereal pictures. Right off the bat, the vehicles are moderately little in measure in vast scale elevated pictures. Also, R-CNNs are intended to identify the jumping box of the objectives without extractin the properties of the question. Thirdly, manual explanation is costly and the accessible manual comment of vehicles for preparing R-CNNs are lacking in nature. To address these issues, a quick and precise veicle detection system is proposed. On one hand to precisely extricate vehicle-like targets, theydeveloped an exact vehicle-proposition network (AVPN) in light of hyperfeature outline joins hierarchial include maps that are more precise for shopping center objet detection. On other hand, they propose a coupled R-CNN strategy, which joins an AVPN and avehicle property learning netork toextract the vehicle's area and qualities at the same time. For unique extensive scale airborne pictures with restricted manual explanations, they utilized trimmed picture blokcs for traiing with information expansion to abstain from overfitting. Complete investigation on the general population Munich Vehicle dataset and a gathered dataset show the precision and adequacy of the proposed technique.

Zhao et al. [5] were inspired by the mechanism of the human vision system and its capability to distinguish nuances processing. It proposes a convolutional neural network (CNN) targeting visual attention classifying images.For generating a focused image, highlighting selected segment of the image while ignoring other details was achieved by visual-attention based image processing module. This helpsto reduce the interferences of the background to the classification. CNN model was inputed the focused image for classification. Reinforced CNN is then implemented using information entropy calculated based on classification probability distribution. This helped the learning agent to achieve more accurate image classification. This is done by an evaluation network. After obtaining the important areas, variousinput components such as vehicle lights, vehicle wiper, etc are extracted from the imge and these feaures are used to detect and locate vehicles. Systemataic experiments on a surveillance-nature datasets which contains images captured by surveillance cameras in the front view are conducted in form of two tasks. The first task ‘Vehicle-5’ is to divide vehicles into 5 caegories according to their types. The second task ‘Vehicle-58’ is vehicle classification discriminated from different vehicle markers. The design requires all images to be sqaures andhence all the images are compressed into squares 224x224. The model proved to be more vesaltile than large-scale CNN.

Xiaozhu et al.[6] implemented a novel mehodto detect armored vehicles in a battlefield. The background becomes complex in a battlefield due to artificial camoflague, varying illumination, shelters, etc. This paper detects armored vehicles using deep learning. The genuine combat zone condition picture is first made into the VOC2007, and the information is prepared by the Faster R-CNN and ZF NET model. The proposed demonstrate utilizes a particular pursuit strategy contrasted with the customary sliding window system of the traditional R-CNN technique. CNN is utilized for highlight extraction and standard database (ImagNet, Pascal VOC) is utilized for preparing and assessment. As indicated by the test results,it is demonstrated that the Faster RCNN ZF display goodly affects the detection and acknowledgment of shielded defensively covered vehicles in front line condition. Contrasted and the conventional method,the question detection and acknowledgment technique in view of the profound learning kills the bulky picture preprocessing link,and its conclusion to-end engineering enormously enhances the detection and acknowledgment effectiveness, and has great versatility,showing a solid application prospect. Furthur inquire about proposed is to build the acknowledgment precision and the exactness of the bouonding box relapse.

Wang et al. [7] found Vehicle-following maps areimportant part of autonomous driving frameworks. Locating and acknowledging presence of tail- light is vital to keep an independent vehicle from backside impacts. Aoustic sonar Advanced Driving Assistance System (ADAS) are available in market for the purpose. Mobile-eyesare also utilized for backside crash cautioning. Researchers have built up a novel approach to deal detecting vehicles and perceiving brake-lights from a solitary picture progressively. Not at all like past methodologies where match taillight must be removed unequivocally, have researchers utilized vehicle raise appearance picture. On an extensive database, "Brake Lights Patterns" (BLP) are learned by a multi-layer observation neural system. Given a picture, the vehicles can be named "brake" or "typical" utilizing the profound classifier. The vehicle can be identified rapidly and vigorously by

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

consolidating multi-layer lidar (IBEO Lux combination framework) and a camera. Street division and a novel vanishing point locale of intrigue (ROI) assurance strategy are investigated to additionally accelerate the identification and enhance the strength. The test comes about directed on some genuine on-street recordings have demonstrated the strength and effectiveness of the proposed approach.

Cui et al. [8] Vehicle re-identification has turned into a basic undertaking in light of the developing blast in the utilization of observation cameras out in the open security. The most generally utilized arrangement depends on tag confirmation. Be that as it may, when confronting the vehicle without a permit, deck autos and other tag data mistake or missing circumstance, vehicle seeking is as yet a testing issue. This paper proposed a vehicle re-distinguishing proof technique in light of profound realizing which abuse a two-branch Multi-DNN Fusion Siamese Neural Network (MFSNN) to wires the order yields of shading, display and stuck blemishes on the windshield and guide them into an Euclidean space where separation can be straightforwardly used to gauge the similitude of discretionary two vehicles. With a specific end goal to accomplish this objective, researchers introduce a technique for vehicle shading recognizable proof in view of Alex net, a strategy for vehicle demonstrate distinguishing proof in view of VGG net, a technique for glued imprints recognition and ID in light of Faster R-CNN. Researchers assess our MFSNN technique on VehicleID dataset and in the analysis. Test comes about demonstrate that our technique can accomplish promising outcome. Table 1 presents a comparison of various vehicle detection and tracking techniques available in literature.

TABLE 1: COMPARISON OF VARIOUS VEHICLE DETECTION TECHNIQUES BASED ON LITERATURE

Citation	Technique used	Parameters	Solution and Accuracy
Tayara et al. [1]	CNN based model is used to regress the vehicle spatial density from an aerial image.	Higher precision and F1 scores compared to conventional technique but requires more time.	Precision:93.30% Recall:90.51%
Chu et al. [2]	Multi-tasking deep CNN model alongside ROI voting	More accurate results on the data tested with the KITTI dataset.	Accuracy:91.67%
Cai et al.[3]	Based on the Bootstrap Aggregating mechanism.Uses deep composite structures for detection.	The proposed method improves the detection rate by an average of approximately 3%.	Detection rate above 92%
Deng et al.[4]	AVPN based on hyper feature map and a coupled R-CNN model	Evaluationon Munich Datasets shows faster and moreefficient performance compared to existing algorithms.	Precision:92%
Zhao et al., [5]	Uses a CNN model that focuses on particular section of image to identify important sections	The proposed model is more competitive compared to large scale CNN models	97.93% test accuracy rate on vehicle-5 dataset.
Xiaozhu et al.6]	Faster RCNN and ZF NET are coupled to detect armored vehicles;	eliminates the cumbersome image preprocessing link, and it's end-to-end architecture greatly improves the detection and recognition efficiency	Detection rate of tank is 97.20% and for wheeled infantry fighting vehicle detection rate is 95.70%
Wang et al., [7]	BLPs are learned by CNN, The vehicle is detected by combining multi-layer lidar and a camera. ROI is used for road segmentation	The average recognition accuracy is 89% and when applied to the vehicle detection, and the average accuracy is 99%	Vehicle-following, obstacle avoidance, etc., have shown that the accuracy and speed of the proposed approaches are satisfied with the navigation requirement.
Cui et al. [8]	A two branch Multi DNN Fusion Siamese Neural Network (MFSNN) is used. It is based on VGG net and	Compared with existed methods, the proposed network structure achieves a high predict accuracy	Accuracy 70.5%

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

	faster RCNN.		
Shi et al. [9]	A selective search method and a target detection model based on Fast R-CNN are used to detect vehicle	The proposed method is superior to the result of multi - feature and classifier detection in terms of accuracy.	Detection rate:86.2%
Bedruz (10)	minimize the effect of occlusion in real time vehicle detection	83 correct prediction out of 102	81% detection rate in high density area
Saini (11)	traffic surveillance for accident detection	It is able to track vehicle even when there are several other vehicles in the camera's field of view.	Maximum number of feature detection in moving vehicle
Sajib (12)	Increase the speed of classification and detection	average detection rate is 97.81% and average classification accuracy is 91.75%	Accuracy:97.81%
Arya (13)	Detect the vehicles using videos of both low and high resolution captured by static camera	Efficiency:85-90%	
GU (14)	Classification and detection of vehicle using CNN	The mAP of our network upto 80.5% in detecting vehicle class	Detection speed is 48 frame per second which can be used for real time vehicle detection
Kim(15)	New licence plate detection method	Precision: 98.39% Recall: 96.84%	Precision:98.39% Recall:96.84%
Zhuang (16)	Rapid vehicle detection using a cascade of strong classifiers	True positive rate: 96.30	Precision:97.2% Recall:95%

Shi et al. [9]As of late the exploration of vehicle location is fundamentally through machine adapting, yet regardless it has low identification exactness issue. With the investigation of analysts, utilizing profound learning strategies for vehicle identification ends up hot. In this paper, a particular hunt technique and an objective recognition demonstrate in light of Fast R-CNN are utilized to recognize vehicle. The system advances the model by preprocessing the example picture and the new system structure. Right off the bat, the trial utilizes general society KITTI informational index and self-gathered BUU-T2Y informational collection, individually, to train approval and test. Also, in view of the first informational index, the trials go ahead through incremental picking up, joining the KITTI dataset with the BUU-T2Y dataset. The test comes about demonstrate that the proposed strategy is better than the consequence of multi - highlight and classifier identification as far as exactness. To an expansive degree, the proposed technique tackled the issue of missing vehicle for recognition and enhanced the precision of vehicle testing and power.

Remaining of the vehicle detection techniques using deep learning approaches can be summarized by following studies. **Bedruz [10]** presented an approach to limit the impact of impediment on managing constant vehicle detection and following. The vehicle detection framework will utilize the detection of vehicles going to theregion of interest (ROI). **Saini[11]** The vehicle detection and following strategy proposed in this paper requires moving highlights to be distinguished and followed additionally utilizing region based square coordinating methodology. **Sajib[12]** proposed focusing on ongoing applications. So the execution time must be quicker in both detection and arrangement. They are proposing a financially savvy and quicker model that can without much of a stretch be executed for ongoing applica-tions. **Arya[13]** aimed to experiment and study continuous vehicle detection and following calculations. It would limit the imperatives like info picture determination, securing stature, natural commotion and so forth on the information document and

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

still ready to accomplish attractive execution. Further, to distinguish the vehicles utilizing recordings of both low and high determination caught by a stationary activity observation camera. **GU et al. [14]** proposed a method of real-time vehicles detection and tracking using Convolutional Neural Networks. They present network architecture, which create multiple vehicle candidates and predict vehicleprobabilities in one evaluation. **Kim [15]** firstly detected vehicle regions with the faster R-CNN algorithm and localise the license plate in each vehicle region with a convolutional neural network (CNN) classifier.

CONCLUSIONS AND FUTURE SCOPE

In the last decade, object detection and tracking remained a problem in research and it has received huge attention across the globe. In the present study, various approaches to detect and track objects in videos are presented. This paper surveys deep learning based vehicle detection techniques used in video processing. It presents a comparative study of the algorithms used in each technique and presents a comprehensive survey on various proposed techniques for vehicle detection. Further, this survey gives a better understanding of the various techiques and highlights the issues and solutions to various challenges faced which can help researchers in choosing an appropriate technique for vehicle detection based on the requirements and environment

REFERENCES

Arya, K. V., Tiwari, S., & Behwalc, S. (2016, June). Real-time vehicle detection and tracking. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2016 13th International Conference on* (pp. 1-6). IEEE.

Bedruz, R. A., Sybingco, E., Quiros, A. R., Uy, A. C., Vicerra, R. R., & Dadios, E. (2016, November). Fuzzy logic based vehicular plate character recognition system using image segmentation and scale-invariant feature transform. In *Region 10 Conference (TENCON), 2016 IEEE* (pp. 676-681). IEEE.

Cai, Y., Wang, H., Zheng, Z., & Sun, X. (2017). Scene-Adaptive Vehicle Detection Algorithm based on a Composite Deep Structure. *IEEE Access*.

Chu, W., Liu, Y., Shen, C., Cai, D., & Hua, X. S. (2018). Multi-task vehicle detection with region-of-interest voting. *IEEE Transactions on Image Processing*, *27*(1), 432-441.

Cui, C., Sang, N., Gao, C., & Zou, L. (2017, November). Vehicle re-identification by fusing multiple deep neural networks. In *Image Processing Theory, Tools and Applications (IPTA), 2017 Seventh International Conference on* (pp. 1-6). IEEE.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 248-255). Ieee.

Deng, Z., Sun, H., Zhou, S., Zhao, J., & Zou, H. (2017). Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(8), 3652-3664.

Gu, L., Jia, D., Vicaire, P., Yan, T., Luo, L., Tirumala, A., & Krogh, B. H. (2005, November). Lightweight detection and classification for wireless sensor networks in realistic environments. In *Proceedings of the 3rd international conference on Embedded networked sensor systems* (pp. 205-217). ACM.

Kim, K. I., Jung, K., & Kim, J. H. (2002). Color texture-based object detection: an application to license plate localization. In *Pattern Recognition with Support Vector Machines* (pp. 293-309). Springer, Berlin, Heidelberg.

Saini, R., Ahmed, A., Dogra, D. P., & Roy, P. P. (2017). Surveillance scene segmentation based on trajectory classification using supervised learning. In *Proceedings of International Conference on Computer Vision and Image Processing* (pp. 261-271). Springer, Singapore.

Sajib, M. S. R., & Tareeq, S. M. (2017, December). A feature based method for real time vehicle detection and classification from on-road videos. In *Computer and Information Technology (ICCIT), 2017 20th International Conference of* (pp. 1-11). IEEE.

Shi, K., Bao, H., & Ma, N. (2017, December). Forward Vehicle Detection Based on Incremental Learning and Fast R-CNN. In *Computational Intelligence and Security (CIS), 2017 13th International Conference on* (pp. 73-76). IEEE.

Tayara, H., Soo, K. G., & Chong, K. T. (2018). Vehicle Detection and Counting in High-Resolution Aerial Images Using Convolutional Regression Neural Network. *IEEE Access*, *6*, 2220-2230.

Wang, J. G., Zhou, L., Pan, Y., Lee, S., Song, Z., Han, B. S., & Saputra, V. B. (2016, June). Appearance-based brake-lights recognition using deep learning and vehicle detection. In *Intelligent Vehicles Symposium (IV), 2016 IEEE* (pp. 815-820). IEEE.

Xiaozhu, X., & Cheng, H. (2017, July). Object detection of armored vehicles based on deep learning in battlefield environment. In *Information Science and Control Engineering (ICISCE), 2017 4th International Conference on* (pp. 1568-1570). IEEE.

Zhao, D., Chen, Y., & Lv, L. (2016). Deep reinforcement learning with visual attention for vehicle classification. *IEEE Transactions on Cognitive and Developmental Systems*.

Zhuang, X., Kang, W., & Wu, Q. (2016). Real-time vehicle detection with foreground-based cascade classifier. *IET Image Processing*, *10*(4), 289-296.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Mei-Hui Peng [1, 2], Bireswar Dutta[1*], Shu-Lung Sun[1]

[1]National Chiao Tung University (NCTU), Institute of Information Management, Hsinchu, Taiwan
[2]Minghsin University of Science and Technology, Hsinchu, Taiwan

clare4260@gmail.com, bdutta67@gmail.com, slsun223388@gmail.com

Correspondence:
Bireswar Dutta
bdutta67@gmail.com

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Abstract

Online shopping has explored significantly worldwide in the last few years. Taiwan is no exception. But although Taiwan has experienced a growth in online shopping, it appears to lag behind the rest of the developed countries in Asia. Study into the factors which influence consumer's online shopping trust and satisfaction is thus imperative in order for Taiwan vendors to develop the appropriate strategies for online trades. This paper pursues to recognize what are the main factors of the consumer's intentions to repurchase online in Taiwan; grounded mainly on theoretical models such as Technology Acceptance Model (TAM) this study specially aims to recognize whether there are relations between consumer's repurchase intention and perceived usefulness, perceived ease of use, perceived privacy, consumers' trust and consumers' satisfaction. A survey was conducted with 225 valid consumers from Taiwan using a structured self-administered questionnaire and data was analyzed using Partial Least Squares (PLS) Structural Equation Modelling (SEM). From the results, we attained new empirical evidence for applying trust and satisfaction for studying consumer repurchases intention. The results indicate that perceived usefulness, and perceived ease of use, employ significant positive influence on consumers' trust, in turn, satisfaction is influenced by perceived usefulness, perceived ease of use, and perceived privacy, which indirectly influences consumers' repurchase intention. Thus, the online retailer must concentrate on the technological characteristics of their shopping website to reduce the privacy concern of consumers regarding the unauthorized access and secondary use of their personal and financial data that can improve consumer trust and satisfaction towards shopping website, and lead to repurchase intention.

Keywords: *Trust, Satisfaction, Repurchase intention, TAM, Structural Equation Modeling (SEM), Taiwan.*

Introduction

The online shopping is one of the types of electronic commerce has flourished rapidly since the middle of the 1990s. The development of online shopping is expected to be accelerated because it has a lot of incentives, for instance, convenience, broader selections Chen et al., [12], competitive pricing, better access to information, product quality, and timely receive product [41]. More than 80% of Internet users all over the world have purchased at least one time from online [20] and the section of the world's population who had purchased products or services from online had increased by approximately 9.21% (from 1.52 billion to 1.66 billion) between 2014 to 2017 [65]. Simultaneous with the organizational interest in online shopping, a significant number of academic researchers are being published related to online shopping [30]. These incidents explain that online shopping has drawn the attention of a great number scholars and practitioners too.

As with any other information system (IS), the success of online shopping depends largely on user satisfaction that will eventually increase consumer repurchase intention. Koufaris [45] noted that passionate and perceptive responses of a consumer's first visit to a web store could influence their intentions to return back and their possibility to make yet inadvertent purchases. In the evaluation of this, online retailers must consider the ways to intensify the level of consumers' satisfaction, will lead to repurchase intention. An encouraging approach can be taken by an electronic retailer is to reduce the uncertainty, which can improve the development and maintenance of consumer-retailer relationships [30].

The primary interface for consumers to purchase product or service from online is the website, a form of information technology (IT). The fundamental significant beliefs of Technology Acceptance Model (TAM) [22], the perceived ease of use (PEOU) and perceived usefulness (PU) have been considered as salient factors for web application acceptance and use. Although TAM initially focused on system usage in the workplace, the development of internet and electronic commerce has created a new perspective within which the models could be tested and recent researches have applied it in online shopping context too [30, 60].

In the offline commerce, the face-to-face interaction may directly satisfy buyers through supporting services. In online shopping, salespeople interact via the interface of the website. So, the challenges facing the online

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

retailers are to ensure the privacy for sensitive contents and transactions. Collier and Bienstock [18] defined privacy as the secrecy of the consumer's information which the online retailer keeps confidential. They also determined that it is the assurance by the online retailer that the personal and financial information of consumers will be protected from potential attack of intruders. According to Bhattacherjee [9], an individual is more likely to intend to undertake continued usage when such usage is perceived to be useful.

Online shopping characteristically involves higher levels of uncertainty than visiting a conventional store because online transactions have a lack of the physical assurances than traditional shopping experiences. Information asymmetry is a drawback in online shopping in which the consumers frequently have incomplete or distorted information about the product [6], the process, the outcome, and the electronic retailer [33]. Prior studies pointed out that trust is fundamental to the online environment [30] and trust is being recognized as a key factor in online shopping [30, 33]. More precisely trust is a vital enabler in relations between geographically disseminated people in the virtual community [36]. According to Kim [42] trust and satisfaction are two stepping stones for successful online shopping relationships. Trust plays a major role in consumer retention, is integrated with the TAM model to explain consumers' behavioral intention [37] and maintaining continuity in buyer-retailer relationships. Lee [46] suggested that satisfaction was one of the important factors explaining the repurchase intention of online consumers. Because, a satisfied consumer is more likely to return to buy.

Acknowledging the importance of electronic commerce' consumer retention, several studies have empirically examined consumer satisfaction, trust, and repurchase intention for electronic commerce services in various countries [7, 11, 23, 55, 56, 72]. Although, the outcomes reported in these studies concerning the influence of trust and satisfaction on consumers' repurchases intention remain inconsistent and inconclusive [16, 23, 56, 72]. Therefore, the purpose of this study is to determine key antecedents that influence consumer satisfaction, trust, and repurchase intention in Taiwan regarding online shopping. Additionally, it aims to determine whether PEOU, PU, and privacy are direct antecedents of both consumer trust and satisfaction, and indirect antecedents of repurchase intention in online shopping.

In what follows, the research questions and the significance of the study are discussed. The research questions addressed in this study are: (i) whether the PEOU, PU, and perceived privacy have a positive effect on consumers trust in online shopping, (ii) whether the PU, PEOU, and perceived privacy are significant factors of consumers satisfaction in online shopping (iii) whether trust and satisfaction significantly effect on consumer repurchase intention.

2. Theoretical Framework

2.1. TAM and trust in online shopping

Trust can be defined as a feeling of security and willingness to depend on someone or something [17]. According to Mayer et al., [49] trust is defined as a belief that the trustee will behave according to the trustor's expectations by showing ability, benevolence, and integrity. Trust has been considered as an important variable in the context of electronic commerce [37]. Generally, trust is related to risk reduction and bridge between two parties. In the online shopping environment, where the element of risk is often higher than in traditional environments, the importance of trust is inflated [32].

TAM has been considered a robust framework to investigate how users develop attitudes towards technology and when they decide to utilize it [22, 39, 45, 67]. Based on the previous literature, trust is a mixed belief-intention variable in trust studies. When trust is integrated into TAM, the trusting intention is replaced by the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

intention variable of TAM. In other words, trust in TAM is trusting belief, reflecting the online consumer perception from online retailers like willing to act matching with consumers' interests, being honest in transaction, don't disclose personal information to others without the consent of users, capable of delivering the offered goods or services as promised[14].

2.2. Perceived Privacy

Perceived privacy is the possibility that online retailers collect data about individuals and use them inappropriately [40]. Largely speaking, privacy has been defined as the right of an individual to be left alone and able to control the release of his or her personal information [71]. In the electronic commerce environment, it is related to a web site's policies on the use of user data [3]. Privacy policies of an online retailer involve the adoption and implementation of a privacy policy, notice, disclosure, and choice/consent of consumers [5]. Benassi [8] states that mechanisms such as trust-providing intermediaries and institutional infrastructures that establish and enforce rules and regulations can build trust by addressing privacy concerns.

Therefore, When Privacy disclosures posted on a web site may reduce a consumer's perceptions of privacy-related risk, result in positive experiences with a firm, and increase the consumer's perceptions that the firm can be trusted [21]. Empirical studies suggested that perceived privacy is a critical factor in consumers' acceptance of online services [37, 62]. Though several studies addressed privacy and trust in the electronic commerce context [24, 51], none of them have included privacy as the major antecedent of trust. Therefore, in this study, we will test whether privacy concerns decrease consumer repurchase behaviors mediating the factors trust and satisfaction.

2.3. Satisfaction

Consumers' satisfaction is a key factor for establishing long-term relationships with them and acquiring their repurchase intentions [46]. Kim [43] stated that because electronic commerce is mainly related to use of a new technological breakthrough, receptiveness to the online environment is important to form a positive relationship with satisfaction. Satisfied consumers are most likely to have the intention to repurchase if the service provider reached or exceeded their expectation [1]. Thus, concerning satisfaction, it can be expected that satisfaction could be a mediator between antecedents (i.e., PEOU, PU, and perceived privacy) and repurchase intention in the online paradigm.

3. Proposed Theoretical Model and Research Hypothesis

3.1. Perceived Usefulness (PU) and Perceived Ease-of-Use (PEOU)

PU is defined as the extent to which an individual believes that using a specific system could improve his or her job functioning, which positively influences on the individual's intention to use that system too [22] and perceived ease of use reflects the difficulty of using online shopping site. Though the concept of trust is the most significant characteristic of electronic retailers, through which consumers react to marketing activities [50]. Braun [10] concluded that trust and usefulness are found to be significant predictors of intention to use social networking websites. If shopping site interface does not have the clear layout and effective navigation, users may sense difficulty to use it. They may also doubt online service retailers' ability and benevolence to provide quality services. Previous research by Pavlou [60] has indicated the influence of perceived ease of use on user trust.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

As an attitudinal variable in TAM, satisfaction plays an intervening role between intentions and PEOU beliefs. Park [58] concluded that satisfaction plays a vital role in determining PEOU and PU of navigation systems. According to Bhattacherjee [9] an individual express positive feelings of satisfaction and believe to use continuously when such exercise is perceived being useful.

H_1: There is a positive relationship between PU and trust.

H_2: There is a positive relationship between PU and satisfaction.

H_3: Perceived ease of use has a positive effect on trust in online shopping.

H_4: Perceived ease of use has a positive effect on consumers' satisfaction in online shopping.

3.3. Perceived Privacy

Privacy echoes the extent to which an online shopping website is safe and protects the consumers' personal and financial information [15]. Flavian [28] showed that consumer distrust rise increasingly concerned as of how personal information has been collected and processed. Consumers who are uncertain whether their privacy is protected will most possibly be reluctant to repurchase from online. On the other hand, if privacy is guaranteed, then they will be more eager to repurchase from online. According to Bart [5] a consumer's feeling of privacy is significant and positively influence on trust in an online shopping. Consumer's uncertainty considering the probability of an online retailer performing dishonestly to them can significantly deteriorate their willingness to participate with the retailer. Tsarenko and Tojib [69] noted that betrayal of personal privacy can lead to defection by consumers.

H_5: There is a significant positive relationship between perceived privacy and trust.

H_6: Perceived level of privacy protection has the significant positive influence on online consumer satisfaction.

3.4. Trust

Trust is a significant antecedent of constructing relationships between buyer and seller. In any buyer–seller relationship, evaluation of consumers' trust before a definite exchange episode is being found to have a direct impact on their post-purchase satisfaction [65]. Harris and Goode [35] suggested that trust is positively and directly associated with satisfaction and this relationship was strongly supported. Gefen [31] showed that consumer trust has a positive effect on the online purchase decision. Weisberg [73] recommended that consumers displayed a higher intention to purchase online in the future when they had higher trust in the online shopping website. Thus, we set forth the following hypotheses.

H_7: Consumers' trust has a positive influence on consumers' satisfaction.

H_8: Consumers' trust in online shopping will positively influence consumers' repurchase intention.

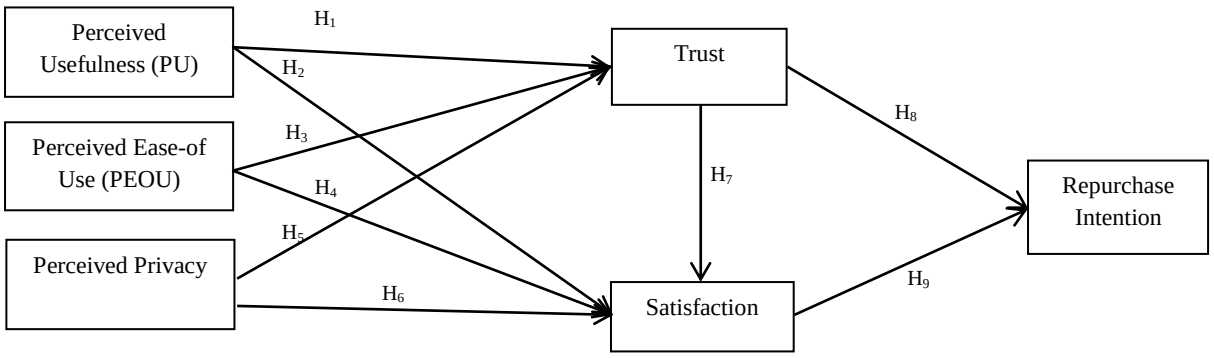


Figure 1: Proposed Theoretical model

3.5. Satisfaction

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

According to Kotler [44] satisfaction is an individual's feeling of desire or dissatisfaction resulting from comparing the perceived performance or outcome of online shopping concerning his or her expectations. Oliver [52] hypothesizes that satisfaction is positively related to future intention, both directly and indirectly through its influence on attitude. Zeithaml [75] recommended that consumers with a higher degree of satisfaction have a tendency of a higher intention to repurchase and recommend the purchased product.

H_9: Satisfaction positively influences repurchase intention.

4 Materials and methods

4.1. Questionnaires design and data collection

A preliminary list of measurement items was initially developed after reviewing the literature regarding TAM, privacy, trust, satisfaction and repurchase intention. Table 4 summarized the lists of items and sources. The study instrument used in this study included three sections. In the first section (cover page), the purpose of the study, and a definition of online shopping was provided. The second section regarded respondents' basic information, including their gender, age, education, and online shopping experience etc. The third section contained indicators regarding PU (5 items), PEOU (3 items), perceived privacy (4 items), trust (4 items), satisfaction (3 items) and repurchase intention (4 items). All the items were measured on a five-point Likert scale, ranging from 1 for strongly disagree to 7 for strongly agree. To enhance the reliability and validity of the indicators, this study modified the content of the items.

Table 1. Measurement Items

Dimension		Items	Sources
Perceived Usefulness (PU)	PU1	Using the online shopping website enables me to finish my shopping tasks more quickly	Davis [22]; Peng et al., [61]; Wen et al., [72]
	PU2	Using the online shopping website for shopping helps me to make better purchase decisions	
	PU3	Using the online shopping website makes it easier to make purchases	
	PU4	Using the online shopping website for shopping helps me to save money	
	PU5	Overall, I find using the online shopping website for shopping is useful	
Perceived Ease of Use (PEOU)	PEOU1	The online shopping website is easy to use	Davis[22]; Peng et al., [61]
	PEOU2	The online shopping website is flexible to interact with online retailer.	
	PEOU3	Finding products from online shopping website is easy.	
Perceived Privacy	PP1	I think the online shopping website abides by personal data protection laws.	Park and Kim [57] ; Peng et al., [61]; Yen and Lu [73]
	PP2	I think online shopping website keeps information about my transactions secret.	
	PP3	I think online shopping website will not share my personal information with others.	
	PP4	I think online shopping website will protect my personal information from unauthorized access.	
Trust	TRU1	I feel safe in my transactions with the online shopping website	Gefen et al. [31]; Peng et al., [61]; Wen et al. [71]
	TRU2	I select online shopping website, which I believe are honest.	
	TRU3	I feel that online shopping website would provide me good service.	
	TRU4	I feel that online shopping website is trustworthy.	

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Satisfaction	SAT1	I was very satisfied with my overall online shopping experience.	Hong et al. [38]; Peng et al., [61]; Wen et al. [71]
	SAT2	I was very pleased with my overall online shopping experience.	
	SAT3	I was absolutely delighted with my overall online shopping experience.	
Repurchase Intention	RPIN1	I intend to continue using online shopping rather than discontinue its use.	Bhattacherjee [9]; Peng et al., [61]; Wen et al., [71]
	RPIN2	My intention is to continue using online shopping rather than use traditional shopping.	
	RPIN3	I would recommend my friends and relatives to use the online shopping website to purchase.	
	RPIN4	If I could, I would like to continue online shopping as much as possible.	

Both a pre-test and a pilot test were conducted to validate the instrument. The pre-test involved six experts, that is, two professors from Information Management (IM), three doctoral scholars with expertise in the electronic commerce field, and one doctoral scholar in the information management field. Respondents were asked to explore the appropriateness of items, the format, and the wording of the scales. The pilot study involved 50 respondents self-selected from the study population. Based on the respondents' reply at the pre-test and pilot test, some items were modified to exhibit the survey's purpose more noticeably. The reliability of all items was acceptable (Cronbach's alpha is above 0.80) and items loaded in confirmatory factor analysis (CFA) are 0.70 or more. Thus, the instrument has validated reliability and content validity. The result of the pilot study is presented in Appendix A (Table A1).

4.3. Research Setting

Cooper and Schindler [19] insisted that convenience sampling is a useful approach during the early stages of the exploratory study. Therefore, this study used a convenience sampling approach to conduct the survey. The purpose in the data collections was to reach as diverse sample as possible that would closely follow the representation of the demographic categories of the general adult individuals and specifically constructed to be representative of the Taiwan population. For that purpose, individuals were approached in various settings, including neighborhoods, small businesses, public meeting places such as parks and transportation stations etc. The individuals were asked to participate completely voluntary and if they wished, a pre-addressed and prepaid envelope was given to them, so they can fill in and return the survey by mail at a later time. A total of 262 questionnaires were given out using the direct procedure, in which 236 respondents replied back. All participants were given consent forms and also informed about their rights to withdraw participation at any time during the study.

4.4. Data Analysis

SPSS and Smart-PLS software, a technique of Structural equation modeling (SEM), were used for statistical analysis. SEM was used for three reasons. First, SEM is a multivariate technique that allows the simultaneous estimation of multiple equations [34]. Second, SEM executes factor analysis and regression analysis in the single step, as SEM is used to test a structural theory. Third, SEM has become a very popular analysis technique in social science researches. All constructs were modeled as reflective, for the model test. Data analysis was conducted on the two-step approach suggested by Anderson and Gerbing [2]. First, testing convergent validity and discriminant validity of the measurement model, and subsequently testing research hypotheses and structural model.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

5. Data analysis and results

5.1 Profile of sample

This study collected 236 responses. Of which eleven of the responses were considered unusable, because of missing data and incomplete answers. This resulted in 225 valid responses for final analysis. Samples included 146 (64.97%) male consumers and 79 (35.03%) female consumers. In terms of age, 175 (77.82%) consumers were aged 18 to 35 and 22.18% consumers were above 35 years old. In terms of education, 152 (67.65%) consumers had a college degree, 25 (11.03%) consumers had Master degree, and only 5 (2.21%) consumers had a secondary grade. Characteristics of respondents are summarized in Table 2.

Table 2 Participants Demographics.

Item	Characteristics	Number	Percentage (%)
Gender	Male	146	64.97
	Female	79	35.03
Age	18-25	83	36.80
	26-35	92	41.02
	36-45	28	12.20
	>=46	22	9.98
Education	Junior high school	5	2.21
	Senior high school	23	10.29
	College	152	67.65
	Master	25	11.03
	Other	20	8.82

5.3. Tests of the measurement model

Reliability analysis was tested using Cronbach's alpha and composite reliability (CR), to measure the model's internal consistency. Table 3 shows the results. The Cronbach's alpha of each construct ranged from 0.908 to 0.977, are above the recommended value of 0.7 by Hair et al., [34]. CR values for the latent factors are above 0.7 suggested by Hair et al., [34] implying good reliability and consistency for the measurement items of each construct.

Table 3 Measurement model.

Constructs	Mean	SD	Item	Loadings	No. of items	Composite Reliability	Standardized Cronbach's α	AVE
PU	4.86	0.96	PU1	0.871	5	0.951	0.934	0.797
			PU2	0.915				
			PU3	0.913				
			PU4	0.843				
			PU5	0.919				
PEOU	4.92	0.90	PEOU1	0.930	3	0.956	0.908	0.879
			PEOU2	0.953				
			PEOU3	0.928				
PP	4.73	0.87	PP1	0.882	4	0.952	0.969	0.834
			PP2	0.930				
			PP3	0.937				
			PP4	0.902				
TRU	4.54	0.93	TRU1	0.953	4	0.979	0.963	0.924
			TRU2	0.969				
			TRU3	0.949				
			TRU4	0.972				
SAT	4.60	0.89	SAT1	0.985	3	0.984	0.956	0.954
			SAT2	0.974				

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

			SAT3	0.970				
RPIN	4.71	0.92	RPIN1	0.970	4	0.983	0.977	0.938
			RPIN2	0.964				
			RPIN3	0.960				
			RPIN4	0.978				

Convergent validity of the scales is examined by using three standards suggested by Bagozzi and Yi [4]: (1) Loadings of each indicator should be higher than 0.7 [27]; (2) CR should be above 0.7; and (3) Average variance extracted (AVE) of each construct should be surpassed the variance because of the measurement error of that construct (i.e. AVE should be exceeded 0.50). As Table 3 confirms, the factor loading of every item in the measuring model of current study exceeded are well above 0.7. CR values are ranged from 0.951 to 0.984. AVE values of constructs are ranged from 0.79 to 0.95, thus meeting each condition for convergent validity.

To test discriminant validity, Fornell and Larcker [27] recommended that the square root of the AVE of the construct should be greater than the estimated correlation shared between the construct and other constructs in the model. Table 4 shows the square root of AVE for each construct was greater than the correlation values of the construct, thus meeting the condition for discriminant validity.

Table 4: Average variance extracted and Discriminant validity.

	PEOU	**PP**	**PU**	**TRU**	**SAT**	**RPINT**	**AVE**
PEOU	**0.937**						0.879
PP	0.463	**0.913**					0.834
PU	0.861	0.516	**0.892**				0.797
TRU	0.784	0.566	0.703	**0.961**			0.924
SAT	0.767	0.624	0.772	0.834	**0.976**		0.954
RPINT	0.664	0.793	0.616	0.814	0.797	**0.968**	0.938

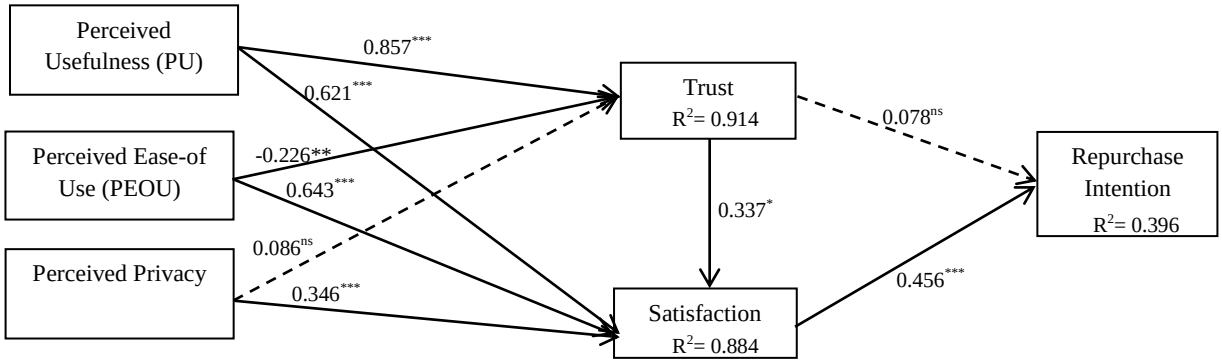
5.4. Tests of the structural model

The summary of the hypotheses test of study model is presented in Table 8. The results provide support for the nine proposed relationships (i.e. H_1, H_2, H_3, H_4, H_6, H_7, H_8, H_9, and H_{11}) while remaining two relationships (i.e. H_5 and H_{10}) is not significant at the 0.05 level of significance. Tests of significance for all the paths are performed by means of the bootstrap resampling procedure. Figure 3 displays the standardized path coefficients, path significances, and variance explained (R^2) by each path. The variance (R-square scores) from the PLS output are as follows: trust 0.914; satisfaction 0.884; repurchase intention 0.396. Thus, the fit of the overall model is good.

Table 5. The Result of Hypotheses testing

Hypothesis	Proposed Hypothesis Relationship	Path Coefficients	*t*-Statistics	Hypothesis Test Results
H_1	PU → Trust	0.857	6.214	Supported
H_2	PU → Satisfaction	0.621	3.741	Supported
H_3	PEOU → Trust	-0.226	2.461	Supported
H_4	PEOU → Satisfaction	0.643	5.142	Supported
H_5	PP → Trust	0.086	0.662	Rejected
H_6	PP → Satisfaction	0.346	3.426	Supported
H_7	TRU → Satisfaction	0.337	2.016	Supported
H_8	TRU → Repurchase Intention	0.078	1.247	Rejected
H_9	Satisfaction → Repurchase Intention	0.456	4.326	Supported

https://sites.google.com/site/ijcsis/
ISSN 1947-5500



Note. * Significant at p<0.05* level, p<0.01**, p<0.001***, ns not significant at p<0.05 level.
Figure 2. Path analysis result

Apart from the evaluation of intrinsic model quality, while explaining the model, it is essential to compare standardized direct, indirect, and total effects of the model before understanding the correlation between the variables [48]. With the respect to key determinants of repurchase intention, consumers' satisfaction has the most direct influence followed by trust. When considering direct and indirect influence on consumers' perceived usefulness exhibits the most influence on repurchase intention followed by satisfaction, perceived privacy, perceived ease of use, and trust (Table 6).

Table 6. Direct and indirect factors predicting trust, satisfaction and repurchase intention

	Trust (TRU)			Satisfaction (SAT)			Repurchase intention (RPIN)		
	D	I	T	D	I	T	D	I	T
PU	0.857		0.857	0.621	0.289	0.91		0.482	0.482
PEOU	-0.226		-0.226	0.643	-0.076	0.567		0.240	0.240
PP	0.086		0.086	0.346	0.028	0.374		0.319	0.319
TRU				0.337		0.337	0.078	0.154	0.232
SAT							0.456		0.456

7. Discussions

The results of this study provide support for the research framework presented in Figure 1 and for the hypotheses regarding the directional relationships among the research constructs. The key contribution of this paper is proposed of an integrated theoretical framework and use of survey data to validate the direct influences of trust and satisfaction on repurchase intention and effects of PU, PEOU, and perceived privacy on repurchase intention mediated by trust and satisfaction. The empirical results indicate that most of the hypotheses of the conceptual model are supported. The findings among the research constructs are discussed below.

Perceived usefulness is a more important factor of trust than perceived ease of use, implying that a more useful online shopping website can encourage consumers to trust in it and simultaneously if the website is easy to navigate will also increase consideration of trust. This finding coincides with the results of previous studies [55, 68]. According to the study by Shin [64] and Li [47] while consumers noticed that an online shopping website is providing quick search, a convenient buying process, speedy access and concise payment, they started to trust it and intend to repurchase from the same shopping website.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Second, PU and PEOU also significantly influence consumer satisfaction. That is, the more useful and easier an online shopping website is, the more satisfactory it is. Consumers are expected to be more satisfied with online shopping website if they consider that using the website will improve their performance, efficiency, and productivity [70]. Thus, website designers should consider an extent of consumer choices and preferences to design a shopping website, which could promote a consumer's satisfaction towards the continuance of online shopping services by concentrating on profitability to have a friendly usage experience and improve consumer's competency and performance. In other words, online retailers must guarantee that consumers' experience during the website visit must satisfy both of their utilitarian purpose (PU and PEOU) and social/psychological factor (satisfaction and trust).

Surprisingly, perceived privacy was not a significant factor of trust. A possible explanation is that experienced internet users are more familiar with privacy technologies. In this way, they can without much of a stretch perceive the highlights, for example, authentications or encryption keys which contributed to privacy criteria of a website. Since these privacy characteristics guarantee almost total privacy, the relative importance of privacy concerns for these users become insignificant. Therefore, the trust in the online shopping website, jointly with the presence of privacy features, drives the decision to disclose personal and financial information with less discomfort. However, this result is supported by prior studies. According to the study by Pavlou and Chellappa [59], the influence of perceived privacy was weak in comparison with the strong influence of perceived privacy on trust. The second possible reason might be that based on the demographic questions of the study, 72.43% consumers have brought products/services 2-15 times from online shopping website. So, the respondents are quite proficient in buying from online shopping websites and well familiar with privacy issues regarding online shopping.

The current study shows that perceived privacy (0.289, t-value = 4.910) consistently has a positive and direct influence on satisfaction, confirms the concerns that consumers often have with the online transaction. A great volume of consumers' information, which is being collected by the online retailer, produces a great extent of privacy concern in consumers. So, ensuring the protection of consumer's privacy contributes significantly to consumer's satisfaction. This finding is consistent with the finding by Tsarenko and Tojib [69], who claimed that disappointed to keep consumer's private information, can lead to a loss of self-confidence and departing of consumers.

Fifth, the study finding confirms that trust drives satisfaction and has a direct impact on consumer satisfaction. Trust is the principal basis for a consumer's acceptance of an online purchase. As discussed prior study trust is one of the significant positive post-adoption beliefs that intensify consumer satisfaction. Consumer satisfaction is also stated as being analogous to a post-consumption behavior when driven by an initial trust which lessens exchange ambiguity. Thus, the current study validates that repurchase intention is well-achieved through the mediator of consumer satisfaction, regarding the proposed antecedents.

Hypothesis 10, (0.097, t-value = 0.556) concerning the negative influence of trust on repurchase intention, was non-significant at 0.05 level. One possible reason could be that trust is less important and typically demonstrates a lower impact in previous research [72], while compared with other factors of behavioral intention [47]. On the basis of the demographic questions of the study, the principal reasons for shopping online are the convenience, better pricing and saving time. Conversely, the reasons for not shopping online are negative concerns associated with shopping online are lack of trust, lack of after sale service and risk concerns. So, the trust could not be the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

main cause that consumers want to continue shopping online. Though, the absence of trust might be the main cause consumers decide not to shop online or why they perceive negative concerns regarding online shopping. That might be another justification for why the direct relationship between trust and repurchase intention was not statistically significant in the current study. The third possible reason might be that trust indirectly affects repurchase intention through satisfaction, and the direct effect is not significant through the mediator.

Sixth, the study finding shows that consumer satisfaction has a strong positive effect on the repurchase intention. This result coincides with the previous study by Oliver and Swan [53, 54]. Consumer satisfaction is mainly based on consumers' previous experience of the transaction with the retailer. The more satisfactory experiences they have, the greater the possibility that consumer will come back to purchase again from the same retailer. Because of the heterogeneous interests of consumers, their satisfaction must be evaluated based on multiple aspects. Thus, online retailers carefully look into the matter of elevating the level of consumer satisfaction, through instance post-service, specific promotions for frequent consumers, etc.

This study contributes to theory and practice in multiple ways. First, incorporated model studied in this study, collective components of the TAM, external variable perceived privacy with trust and satisfaction have overcome the limited applicability of the TAM to study users' intention to repurchase and the results of the study improve the current understanding in the field of technology acceptance and electronic commerce implementation. Second, the study instrument offers not only an inclusive evaluation but also has the capability to consider what traits of the repurchase intention (technology, behavioral or user's demographic differences) are challenging from the users' standpoint. Third, our development of the TAM model explicates how differences in usage intention are affected by the perception of repurchase intention in users. The acceptance theory developed from the current study might be enhanced for application in large-scale services and organizations considering the repurchase intention. Finally, the results of this study initiate better technology usage and could also have a better consideration for electronic commerce providers and policymakers before taking the decision about further spending on new IS implementation.

8. Conclusions

This study is expected to give contribution in form of better understanding and explanation of on trust, satisfaction and repurchase intention from online shopping perspective among Taiwan consumers. A theoretical research framework which hypothesizes the key factors affecting three constructs is developed and statistically validated. The significant key factors are: factors related to technological acceptance (i.e., PU and PEOU), and privacy concern. Both PU and PEOU were found to have a significant positive influence on consumer trust and satisfaction. It is also found that privacy concerned is strongly related to consumer satisfaction but weakly related to consumer trust.

This emphasizes that convenient to find the information, quality of information, service quality and easy payment process is found to be indirectly influencing on Taiwan online consumer's repurchase intention. Moreover, the perception of privacy, as perceived by online consumers, is a key issue for the development of online consumer satisfaction. To conclude, while the perception of privacy is high and the transactional relationship is long (online purchasing), trust plays as a key factor of behavioral intention, hereafter; users are likely to provide more personal and financial information with a lesser level of concern. Thus, online retailers must concentrate to improve the privacy of the target system and privacy features should be considered as a principal concern during the shopping website system's design. As users are more satisfactorily inclined toward

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

using the shopping website when they perceive that their transaction process is secure and third parties will not have a possibility to access it.

This study also found that consumer satisfaction significantly impacts on consumer repurchase intention. However, consumer trust has the indirect significant effect on repurchase intention through consumer satisfaction but direct effect is not significant. Repurchase intention is relatively a more deliberative decision, concerning more about overall experience gained from previous transactions and this experience increases the possibility of purchasing again. Additionally, consumers may tend to stay on the same online retailer if they trust the service provider. In other words, consumer's trust in the particular online retailer is likely to enhance the retailer's reputation, leading to repeat purchase. But unclear information, poor service may lead to consumer's distrust of that retailer and switching to another retailer. Therefore, we recommend that online retailer must need to focus on consumer's online shopping experience and should devote effort to improve consumer trust through effective searching policies, convenient payment mechanisms, ensuring transaction security, user-friendly interface, and providing complete information regarding products.

9. Limitation and future research

We note that our findings must be interpreted in view of the study's limitation. First, respondents answered the questions based on their experience of various online shopping websites, instead of responding to questions about a specific website, but the business nature of the website (B2C or consumer-to-consumer (C2C)), and the distinctive kind of designs which may influence on consumers' experience and perceptions of online shopping. So, further research might probably consider the influence of various types of shopping websites and the effect of different website designs. Second, the results may have been affected by self-selection bias. Because of, our sample consists of only active consumers. Consumers who had already stopped to purchase the product from online shopping website may have different perceptions about the effect of TAM variables, privacy, and so possibly will have been differently influenced by them. So, additional research is required for non-active consumers or non-satisfied consumers.

References

[1] Alam, S.; Yasin, N. An investigation into the antecedents of customer satisfaction of online shopping. Journal of Marketing Development and Competitiveness 2010, 5, 1, 71‐78.
[2] Anderson, J.C.; Gerbing, D.W. Structural equation modeling in practice: A review and recommended two step approach. Psychological Bulletin 1998, 103, 411–423.
[3] Angriawan, A.; Thakur, R. A Parsimonious Model of the Antecedents and Consequence of Online Trust: An Uncertainty Perspective. Journal of Internet Commerce 2008, 7, 1, 74-94.
[4] Bagozzi, R.P.; Yi, Y. On the evaluation of structural equation models. Journal of Academy of Marking Science 1988, 16, 74–94.
[5] Bart, Y.; Shankar, V.; Sultan, F.; Urban, G.L. Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study. Journal of Marketing 2005, 69,133-152.
[6] Ba, S.; Pavlou, P.A. Evidence of the effect of trust building technology in electronic markets, price premiums, and buyer behavior. MIS Quarterly 2002, 26, 243–268.
[7] 22. Balla, B.E.; Ibrahim, S.B.; Ali, A.H. The Impact of Relationship Quality on Repurchase Intention Towards The Customers of Automotive Companies in Sudan. British Journal of Marketing Studies 2015, 3, 4, 1-15.
[8] Benassi, P. TRUSTe: an online privacy seal program. Communications of the ACM 1999, 42, 2, 56–57.
[9] Bhattacherjee, A. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 2001, 25, 3, 351-370.
[10] Braun M-T (2013) Obstacles to social networking website use among older adults. Computers in Human Behavior, 29(3):673-680

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[11] Bulut, Z.A. Determinants of Repurchase Intention in Online Shopping: a Turkish Consumer's Perspective, International Journal of Business and Social Science Vol. 6, No. 10; October 2015
[12] Chen, L.; Gillenson, M.L.; Sherrell, D.L. Enticing online consumers: an extended technology acceptance perspective. Information and Management 2002, 39,705–719.
[13] Chen, Y.T.; Chou, T.Y. Exploring the continuance intentions of consumers for B2C online Shopping Perspectives of fairness and trust. Online Information Review 2012, 36, 1, 104-125.
[14] Chi, W. H., & Tang, T. W. (2005). The role of trust in customer online shopping behavior: perspective of technology acceptance model. http://www.casos.cs.cmu.edu/events/conferences/2005/2005_proceedings/Tang.pdf
[15] Chiu, C.M.; Chang, C.C.; Cheng, H.L.; Fang, Y.H. Determinants of customer repurchase intention in online shopping. Online Information Review 2009, 33, 4, 761-784.
[16] Choi, S.A., and Park, J.W., 2015. Investigating the effect of online service quality of internet duty-free shops on trust and behavioral intention. Journal of Airline and Airport Management, 5(2), pp.101-115.
[17] Chung, N.; Kwon, S.J. Effect of trust level on mobile banking satisfaction: a multi-group analysis of information system success instruments. Behaviour & Information Technology 2009, 28, 6, 549–562.
[18] Collier, J.; Bienstock, C. How do customers judge quality in an e-tailer? MIT Sloan Management Review 2006a, 48, 1, 35-40.
[19] Cooper, D. R.; Schindler, P. S. Business research methods (Seventh edition). Irwin/McGraw-Hill: New York. 2001,192.
[20] Cpcstrategy Blog. http://www.cpcstrategy.com/blog/2013/08/ecommerce-infographic/. 2013.
[21] Culnan, M.J.; Armstrong, P.K. Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation. Organization Science 1999, 10, 1, 104–115.
[22] Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989, 13, 3, 319–340.
[23] Dutta, B. Exploring the Factors of Consumer Repurchase Intention in Online Shopping. International Journal of Computer Science and Information Security 2016, 14, 12.
[24] E. Kim, A Model of sustainable trust in B2C e-markets, in: Proceedings of the Seventh Americas Conference on Information Systems, Boston, MA, 2001, pp. 804–809
[25] E.W.T. Nagai, F.K.T. Wat, A literature review and classification of electronic commerce research, Information & Management 39 (5), 2002, pp. 415–429.
[26] Flavia´n, C.; Guinalı´u, M. Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. Industrial Management & Data Systems 2006, 106, 5, 601-20.
[27] Fornell, C.; Larcker, D.F. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 1981, 18, 39–50.
[28] Flavian, C.; Casalo, L.; Guinaliu, M. The role played by perceived usability, satisfaction, consumer trust on website loyalty. Information and Management 2006, 43, 1, 1–14.
[29] Gefen, D.; Karahanna, E.; Straub, D.W. Potential and repeat e-consumers: the role of and trust vis-a`-vis TAM. IEEE Transactions on Engineering Management 2003a, 50, 3, 307-21.
[30] Gefen, D.; Straub, D. Managing user trust in B2C e-services. e-Service Journal 2003, 2, 2, 7-24.
[31] Gefen, D.; Straub, D.W.; Boudreau, M.C. Structural equation modeling and regression: guidelines for research practice. Communications of the Association for Information Systems 2000, 4, 7, 1–80.
[32] Grabner, K.S.; Kaluscha, E.A. Empirical research in on-line trust: a review and critical assessment. International Journal of Human-Computer Studies 2003, 58, 783-812.
[33] Grabner-Kraeuter, S. The role of consumers' trust in online-shopping. Journal of Business Ethics 2002, 39, 43–50.
[34] Hair, J.F.; Anderson, R.E.; Tatham, R.L.; Black, W. Multivariate Data Analysis. NJ: Prentice-Hall, Inc. 1998.
[35] Harris, L.C.; Goode, M.H. The four levels of loyalty and the pivotal role of trust: a study of online service dynamics. Journal of Retailing 2004, 80, 2, 139–158.
[36] Heijden, H. Vd.; Verhagen, T.; Creemers, M. Understanding online purchase intentions: contributions from technology and trust perspectives. European Journal of Information Systems 2003, 12, 1, 41-8.
[37] Hoffman, D.L.; Novak, T.P.; Peralta, M.A. Building consumer trust in online environments: the case for information privacy. Communications of the ACM 1999, 40, 4, 80-5.
[38] Hong, S.J., Thong, J. and Tam, K.Y. (2006), "Understanding continued information technology usage behavior: a comparison of three models in the context of mobile internet", Decision Support Systems, Vol. 42, pp. 1819-1834.
[39] Hu, P.J.; Chau, P.Y.K.; Liu Sheng, O.R.; Tam, K.Y. Examining the Technology Acceptance Model Using Physician Acceptance of Telemedicine Technology. Journal of Management Information Systems 1999, 16, 2, 91-112.
[40] Jarvenpaa, S.L.; Toad, P.A. Consumer reactions to electronic shopping on the World Wide Web.

International Journal of Electronic Commerce 1996, 1, 2, 59-88.
[41] Keeney, R.-L. The value of internet commerce to the customer. Management Science 1999, 45, 4, 533–542.
[42] Kim, D.J.; Ferrin, D.L.; Rao, H.R. Trust and Satisfaction, Two Stepping Stones for Successful E-Commerce Relationships: A Longitudinal Exploration. Information Systems Research 2009, 20, 2, 237-257.
[43] Kim, W.; Lee, Y.; Yoo, Y. Predictors of relationship quality and relationship outcomes in luxury restaurants. Journal of Hospitality & Tourism Research 2006, 30, 2, 143-169.
[44] Kotler, P. Marketing Management. The Millennium Edition. Prentice Hall of India, 2000.
[45] Koufaris, M. Applying the technology acceptance model and flow theory to online consumer behavior. Information Systems Research 2002, 13, 2, 205-223.
[46] Lee, H.; Choi, S.Y.; Kang, Y.S. Formation of E-Satisfaction and Repurchase Intention: Moderating Roles of Computer Self-Efficacy and Computer Anxiety. Expert Systems with Applications 2009, 36, 7848-7859.
[47] Li, D.; Browne, G.J.; Wetherbe, J.C. Why do internet users stick with a specific web site? A relationship perspective, International Journal of Electronic Commerce 2006, 10, 4, 105-141.
[48] Lin, K.Y.; Lu, H.P. Why people use social networking sites: An empirical study integrating network externalities and motivation theory, Computers in Human Behavior 2011, 27, 1152-1161
[49] Mayer, R.; Davis, J.; Shoorman, F. An Integrative Model of Organizational Trust. The Academy of Management Review 1995, 20, 3, 709-734.
[50] Morgan, R.; Hunt, S. The commitment-trust theory of relationship marketing. Journal of Marketing 1994, 58, 3, 20-38.
[51] Nagai, E.W. T., Wat, F. K. T.,"Human resource information systems: a review and empirical analysis, Personnel Review, 35(3), pp. 297-314, 2006.
[52] Oliver, R.L. A cognitive model for the antecedents and consequences of satisfaction. Journal of Marketing Research 1980, 17, 4, 460-469.
[53] Oliver, R.; Swan, J. Consumer perceptions of interpersonal equity and satisfaction in transactions: a field survey approach. The Journal of Marketing 1989a, 53, 2, 21-35.
[54] Oliver, R.L., Swan, J.E. Equity and disconfirmation perceptions as influences on merchant and product satisfaction. Journal of Consumer Research 1989b, 16, 3, 372-83.
[55] Pahlevani, S. Investigating Factors Affecting the Online Repurchasing Intention, Trends in Life Sciences 2015, 4, 1, 197-202.
[56] Pappas, Ilias, Mikalef, Patrick, Giannakos, Michail N., Twenty-Fifth European Conference on Information Systems (ECIS), Guimarães,Portugal, 2017, Value Co-Creation And Trust In Social Commerce: An FSQCA Approach
[57] Park, C.; Kim, Y. Identifying Key Factors Affecting Consumer Purchase Behavior in an Online Shopping Context. International Journal of Retail & Distribution Management 2003, 31, 1, 16–29.
[58] Park, S. Y., Nam, M. W. & Cha, S. B. (2012). University students' behavioral intention to use mobile learning: Evaluating the technology acceptance model. British Journal of Educational Technology.
[59] Pavlou PA, Chellappa RK (2001) The role of perceived privacy and perceived security in the development of trust in electronic commerce transactions. working paper, eBizLab, Marshall School of Business, University of Southern California, Los Angeles, CA.
[60] Pavlou, P.-A. Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. International Journal of Electronic Commerce 2003, 7, 3, 69–103.
[61] Peng, M.-H, Dutta, B., Sun, S.-L. Understanding Factors Influencing Online Shopping Value and Consumer Repurchase Intention: An Exploratory Study in Taiwan Perspective, International Conference on Modern Management and Innovation (2018 ICMMI), Hsinchu, Taiwan, April 27, 2018.
[62] Poon, W. C. (2008). Users' Adoption of E-banking Services: The Malaysian Perspective. Journal of Business and Industrial Marketing, 23 (1), pp. 59–69.
[63] Rezaei, S.; Amin, M. Exploring online repurchase behavioural intention of university students in Malaysia. J. Global Business Advancement 2013, 6, 2.
[64] Shin, K.Y.; Choo, G.W.; Park, T.J. Determinants of Using Internet shopping malls. Journal of MIS Research 2001, 10, 1, 279-301.
[65] Singh, J.; Sirdeshmukh, D. Agency and trust mechanisms in consumer satisfaction and loyalty judgments. Journal of the Academy of Marketing Science 2000, 28, 1, 150–167.
[66] Statista. https://www.statista.com/statistics/251666/number-of-digital-buyers-worldwide/. 2017.
[67] Straub, D. W., Limayem, M., and Karahanna, E. "Measuring System Usage:Implications for IS Theory Testing," Management Science (41:8), 1995, pp.1328-1342.
[68] Suwunniponth, W. Factor Driving Consumer Intention in Online Shopping, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering 2014, 8, 6, 1949-1953.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[69] Tsarenko, Y.; Tojib, D.R. Examining customer privacy concerns in dealings with financial institutions. Journal of Consumer Marketing 2009, 26, 7, 468-476.
[70] Wang, W.; Hsieh, P.; Butler, J.; Hsu, S.H. Innovate with complex information technologies: a theoretical model and empirical examination. Journal of Computer Information Systems 2008, 49, 1, 27-36.
[71] Warren, S.D.; Brandeis, L.D. The right to privacy. Harvard Law Review 1890, 4, 5, 193–220.
[72] Wen, C.; Victor, R.P.; XU, C. An Integrated Model for Customer Online Repurchase Intention. Journal of Computer Information Systems 2011, 52, 14-23.
[73] Weisberg, J.; Te'eni, D.; Arman, L. Past purchase and intention to purchase in e-commerce: The mediation of social presence and trust. Internet Research 2011, 21, 1, 82-96.
[74] Yen, C.H.; Lu, H.P. Effects of e-service quality on loyalty intention: an empirical study in online auction. Managing Service Quality 2008, 18, 2, 127-146.
[75] Zeithaml, V.A.; Berry, L.L.; Parasuraman, A. The behavioral consequences of service quality. Journal of Marketing 1996, 60, 2, 31–46.

Appendix A

Table A1 Results of confirmatory factor analysis and reliability analysis

Constructs	Item	Loadings	Standardized Cronbach's α
Perceived Usefulness (PU)	PU1	0.951	0.969
	PU2	0.958	
	PU3	0.970	
	PU4	0.951	
	PU5	0.962	
Perceived ease of Ease (PEOU)	PEOU1	0.982	0.961
	PEOU2	0.975	
	PEOU3	0.934	
Perceived Privacy (PP)	PP1	0.981	0.988
	PP2	0.982	
	PP3	0.979	
	PP4	0.989	
Trust (TRU)	TRU1	0.865	0.974
	TRU2	0.981	
	TRU3	0.968	
	TRU4	0.961	
Satisfaction (SAT)	SAT1	0.959	0.967
	SAT2	0.945	
	SAT3	0.957	
Repurchase intention (RPIN)	RPIN1	0.954	0.952
	RPIN2	0.935	
	RPIN3	0.951	
	RPIN4	0.901	

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Distributed nonhierarchical PKI based on NTRU for secure routing in MANETs

Alaa Moualla
Department of Telecommunication, Higher Institute for Applied Sciences and Technology,
Damascus-SYRIA
alaa.moualla@hiast.edu.sy

Oumayma Al Dakkak
Department of Telecommunication, Higher Institute for Applied Sciences and Technology,
Damascus-SYRIA
oumayma.dakkak@hiast.edu.sy

Mohamad Aljnidi
Department of Informatics, Higher Institute for Applied Sciences and Technology,
Damascus-SYRIA
mohamed.aljnidi@hiast.edu.sy

***Abstract*— A Mobile Ad hoc Network (MANET) is a network of wireless mobile devices deployed without any pre-existing infrastructure or centralized administration. This technology is studied with a number of serious challenges that need to be solved before its successful deployment. An appreciable number of routing protocols used in MANET have left the critical aspect of security out of consideration. In this paper, we remind some of the on-demand routing protocols used in MANET. Further, we propose a solution to increase security for AODV[1], through authenticating nodes without the need of a trusted third party (TTP). This solution depends on a public key infrastructure (PKI) based on NTRU[2] algorithm. AVISPA[3] tool will be used to test this security.**

Generally, public key infrastructures are assumed unavailable in MANETs due to its non-centralized administration. Therefore, we propose a distributed nonhierarchical PKI model based on NTRU algorithm to overcome these challenges. The key element of our approach is to use Lattice based cryptography, specifically NTRU, and a distributed PKI model. We show that the proposed scheme is resistant to a wide range of security attacks and can scale easily to large-size networks.

Keywords - MANETs, ad-hoc routing protocol, AODV, NTRU, PKI, AVISPA, certificate authority (CA), trusted third party (TTP), Threshold-based Cryptography

I. INTRODUCTION

A mobile ad hoc network (MANET) consists of nodes that communicate with each other via wireless media. MANETs do not have any centralized administration nor do they require any fixed network infrastructure and this facilitates quick set up.

MANET has several characteristics, which make its deployment and management very challenging. These include: dynamic topology, constrained resources, lack of infrastructure, shared wireless medium, limited bandwidth and distributed operations.

Routing in MANETs is achieved via the use of a number of -widely used- specific protocols that attempt to counter the challenge posed by dynamic topologies. However, ad hoc network routing protocols have generally assumed an environment where all the nodes are co-operative and trustworthy; hence no security mechanism has been considered [1]. In this paper, we consider only the Ad hoc On-demand Distance Vector (AODV) routing protocol for its properties [2]: dynamic nature, better performance with minimum overhead, in addition to its wide spread acceptance.

Providing security in MANETs is an inherently challenging problem due to the lack of a fixed infrastructure, the dynamically changing network topology, the limitations of the wireless channel, and the limited capabilities of the nodes.

MANETs cannot always guarantee online access to a centralized Certificate Authority (CA) due to the often intermittent and unreliable nature of the wireless channel. Thus, the use of a standard public key infrastructure (PKI) is generally impractical in an ad hoc wireless environment.

Many researches tried to avoid the need of CA in MANET by carrying out authentication based on a shared secret or password established prior to the deployment of the network. A security scheme similar to Pretty Good Privacy (PGP) has been proposed for MANETs [3], whereby certificates are issued by users based on the establishment of chains of trust. This approach is well suited to ad hoc networking environment, but provides only probabilistic security guarantees and relies on transitive trust relationships, which may not be sufficient for some applications. Another approach to securing MANETs is based on a distributed certification authority [4] to avoid the problem of a single point of failure found in traditional PKI architectures. This approach provides deterministic security

[1] Ad hoc On-Demand Distance Vector (AODV) Routing is a routing protocol for mobile ad hoc networks (MANETs) and other wireless ad hoc networks.
[2] NTRU – Nth Degree Truncated Polynomial Ring Units.
[3] AVISPA stands for *Automated Validation of Internet Security Protocols and Applications*.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

guarantees, but raises critical issues of scalability in practical MANETs.

We propose a comprehensive approach to providing a distributed CA-based PKI in MANETs, which will be used in AODV to secure the routing via authentication using NTRU algorithm. The key elements of our approach are: (1) a scheme for dynamically partitioning the network into smaller clusters of nodes based on nodal mobility, (2) a distributed certification authority with multiple CA servers, (3) use of NTRU cryptography.

The distribution of the key management architecture over clusters reduces the storage requirements at the nodes and the CA servers. It also reduces the computational and signaling overhead. Finally, the use of NTRU cryptography dramatically reduces the computations involved in cryptographic operations, making the PKI-based scheme feasible even for nodes of modest computational power, and reduces the need of CA to face the vulnerabilities of MANET routing protocols.

The rest of paper is organized as follows: section 2 presents the related works. Section 3 discusses the vulnerabilities in MANETs and the secure routing protocols. Section 4 presents the proposed PKI model based on NTRU. The analysis is detailed in section 5, and the simulation results using AVISPA are given in section 6. Finally, the conclusion is given in section 7.

II. Related Works

Most of the studies until now are about secure data transmission in MANET [5].

Ashish Sharma et al [6] have focused only on active attacks in the network layer. AODV is a reactive routing protocol for ad hoc networks that maintains route only between nodes that want to communicate, using routing messages. Secure-AODV (SAODV) is a secure routing protocol based on trust model for mobile ad hoc network. To provide security and increase performance in MANET, they have applied SAODV protocol using Hybrid Cryptography Technique (DES, RSA Algorithms) on it.

Kumar Verma et al [7] present a robust and secure mechanism for nodes authentication in MANETs. The proposed authentication mechanism is based on certificate exchange between the nodes. It uses digital signature with a hash function to maintain the authenticity of certificates. Simulation for this mechanism shows better performance in terms of throughput, end-to-end delay and packet dropping, in presence of malicious nodes in the MANET.

Jasdanwala et al [8] have proposed just end-to-end data transmission security using NTRU. They also assumed a clustered MANET environment; i.e. the mobile nodes have an efficient cluster-based routing algorithm deployed, which gives the list of clusters available in the environment along with its cluster heads and competent routing between them.

Other studies focus on secure routing protocol against specific attacks like [9], where Tan et al use cryptography technique for secure AODV routing protocol and data transmission against BlackHole attack.

While Sajyth et al [10] have used a Bee clustering approach to apply Elliptical Curve cryptography (ECC) on secure communication-AODV (SC-AODV), provide efficient energy consumption, and secure data delivery system.

Veerpal Kaur et al [11] Have proposed a defense hybrid and clustering mechanism against multiple black hole nodes in a MANET.

Kumar Verma et al [12] have presented an authentication protocol for MANET, which is based on certificate exchange between nodes by using digital signature with a hash function to maintain the authenticity of certificates.

III. Vulnerabilities in MANETs

Security service requirements in MANETs are similar to wired or any wireless network. There are five major security goals, which are needed to protect the data and resources from attacks [13]: authentication, availability, confidentiality, integrity and non-repudiation.

Attacks on MANET's are divided into two major categories: passive and active [14].

- Passive attacks are trying to steal the vital information inside data packets. Examples of passive attacks in ad hoc networks are eavesdropping attacks and traffic analysis attacks.
- Active attacks are trying to interrupt the functionality of the network through reading and changing the information on the data packets, denial of service, altering the routing path by changing routing information … etc.

In the following sub-paragraphs, some MANETs' attacks will be enumerated followed by some proposed methods –in the literature- to secure on-demand routing [15]:

A. Attacks on MANETs

- **Attacks using modification**

The attacker can make some changes to the routing messages, like modifying the route sequence number, or the hop count.

- **Impersonation attacks**

A malicious node changes its identity (such as IP address or MAC address) to an authorized node in the outgoing packets, so it can isolate the authorized nodes from the network or simply change the network topology.

- **Attacks using fabrication**

On this attack, malicious nodes can fabricate their own packets to cause confusion in the network operations.

- **Rushing attacks**

In on-demand routing protocols only one route request packet is forwarded to find the path to the destination. This property is abused in rushing attack by forwarding more request packets.

In the following sub paragraph, we discuss the methods in the literature to secure on-demand routing protocols, taken into consideration in the present paper.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

B. Methods proposed to secure on-demand routing protocols

AODV is the most popular on-demand ad-hoc routing protocol. It enables nodes to pass messages to nodes that cannot directly reach via other neighbor nodes. This is achieved by using a routing table for each node in the network. AODV is able to handle changes in routes and can create new routes in case of an error or a change in topology [16].

There are many techniques proposed for facing the vulnerabilities encountered in on-demand protocols. We mention some of them [16]:

- **Secure AODV (SAODV)**

It is an extension of the AODV protocol. Routing messages in SAODV are authenticated to guarantee their integrity. The source node signs the routing message with its private key, and the recipient nodes verify the signature using the public key of the source. A hash chain mechanism is used to prevent any modification or tampering of the hop count. In fact, the sender cannot sign the hop count because it is incremented at each hop.

SAODV suffers from performance issues caused by the heavy computation of asymmetric cryptographic methods because every node must generate or verify a signature each time it generates or receives a routing message respectively.

- **Security-Aware Ad hoc Routing (SAR)**

SAR attempts to implement a more generalized way of providing security to routing protocols by incorporating security metrics into its base routing protocols and alters its forwarding behavior.

SAR prevents a few attacks such as spoofing and the black hole attack, but it is vulnerable to DoS (Denial of Service), Wormholes and Rushing.

- **Secure Routing Protocol (SRP)**

According to this protocol, the two communicating nodes must have a shared key for communication and verification. SRP was claimed to guarantee the acquisition of accurate topological information; a node initiating the route discovery process is able to discard replies from malicious nodes claiming false topological information, thus ensuring maximum safety.

This protocol can prevent the black hole attack and the attacks due to incorrect routing information, but it is vulnerable to the wormhole, rushing, DoS (Denial of Service) and invisible node attacks.

- **Ariadne**

Ariadne is a robust protocol based on Dynamic Source. It makes use of symmetric key cryptography. It also uses a one way hash along with a message authentication code (MAC) using a shared key between the source and the destination in order to authenticate the source at the destination.

Ariadne is vulnerable to the invisible node attack, rushing, wormhole and DoS (Denial of Service) attacks.

- **Authenticated Routing for Ad hoc Networks (ARAN)**

ARAN is a secure routing protocol, that makes use of asymmetric cryptography and all nodes must request a certificate from the CA when they want to join the network.

ARAN uses peers and trusted third parties to ensure safety in ad hoc networks. The five major components in ARAN are: Certification, Authenticated Route Discovery, Authenticated Route Set up, Route Maintenance, and Key Revocation.

In ARAN, impersonation is easily avoided, because of authentication in every hop. However, since it uses asymmetric key cryptography, the overall performance will slow down. ARAN is also affected by the Wormhole, black hole, DoS and the rushing attacks.

We believe, NTRU cryptosystem can present a smart solution to some of the above-mentioned methods' shortcomings. We will rely on ARAN algorithm with NTRU cryptography to overcome these vulnerabilities. This solution has two main advantages: first, the simple computation of NTRU, which improves the performance. Second, there is no need for a CA anymore, so different security solutions may be achieved while keeping the decentralized nature of MANETs maintained.

IV. PROPOSED METHOD

We need a method to make the AODV routing protocol secured, so it can detect malicious actions and afford protection against them. To do so, one can require it to be authenticated through using a TTP (Trusted Third Party). However, this TTP may facilitate some attacks (man in the middle, brute force …) against the network. The main goal of our method is to use NTRU in authentication, so we do not need a TTP and we can depend on other nodes to give certificates to new nodes in the network. This means the hierarchical structure of the traditional PKI will not exist anymore and the need of CA will disappear.

To overcome the TTP problem, Huei Lu et al [17] extend models already proposed by Shamir and Okamoto [18] to apply mutual authentication and key exchange for MANETs. We use this authenticated method to overcome the need of TTP in AODV. Moreover, we use authenticated routing protocol for initiating the route. The proposed protocol consists of three main stages: the registration process, the route initiation and the route maintenance. These steps are described as follows:

A. Registeration

A special node called key generator (KG), or "dealer" -as in threshold cryptography-, emulating the CA in PKI, is needed in this phase to verify the set of nodes, and to generate the keys. After the registration of all nodes, the KG can be stopped or put off-line, and the other nodes concatenate a key used to authenticate the new nodes that want to join the network.

During registration phase, the KG randomly chooses two polynomials $f \in L_f, g \in L_g$, where L_f, L_g are set of polynomials of degree (N-1) with integer coefficients, and compute the public key

$$K_{ca+} = f_q^{-1} \times g \pmod q$$

Any node A uses a personal identification number ID_A to register to KG.

For each node A, KG computes the public key

$$K_{A+} = f_{A,q}^{-1} \times g \pmod q$$

And the private key

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$K_{A-} = \left(f_A, f_{A,q}^{-1}, f_{A,p}^{-1}, g\right)$$

Then it randomly chooses a polynomial $k_A \in L_k$, where also L_k is a set of polynomials of degree (N-1) with integer coefficients, and computes

$$X_A = k_A \times K_{ca+} \pmod q$$
$$HID_A = H\left(K_{A+} \| ID_A\right)$$
$$S_A = HID_A \times K_{A+} + k_A \pmod q$$

Where H denotes the hashing function and "||" denotes concatenation.

When a new node wants to join the network, it can join easily using threshold cryptography, without need of KG, just a group of nodes will concatenate the key and give it to that new node after they authenticate it. We are working on this issue, and we will describe it in a future paper.

B. Route Initiation

KG is available only at the network initiation time, so each node A contacts CA, which is KG in this phase, to request a certificate including its address and its public key.

$$CA \rightarrow A : cert_A = \left[IP_A, K_{A+}, t, e\right]K_{CA-}$$

The certificate contains the IP address of the node, the node's public key, a timestamp showing the certificate creation date (t) and its expiration date (e). These variables are concatenated and signed by the private key of KG.

Now each node M knows the public key of KG. Once all the nodes have registered and got their keys with the variables X_M, HID_M, S_M the KG has no need to exist in the network anymore.

To initiate the route, each node must authenticate its neighbors before starting route discovery. This authentication will be held in two steps: computation of envelope key and the mutual authentication. In the following, we develop these two sub-steps, before describing the route discovery.

B.1 Computation of the envelope key

To start routing, each node A must communicate securely with its neighbor (or one of its direct neighbors) B, so they should generate the envelope key. Before this, the two nodes need to verify whether $\left(ID_A, H\left(X_A\right), S_A, K_{A+}\right)$ and $\left(ID_B, H\left(X_B\right), S_B, K_{B+}\right)$ are sent from A, B respectively. To do this:

- A sends $\left(ID_A, H\left(X_A\right), S_A, K_{A+}\right)$ to B.
- B sends $\left(ID_B, H\left(X_B\right), S_B, K_{B+}\right)$ to A.
- A computes:

$$\overline{HID_B} = H\left(K_{B+} \| ID_B\right)$$
$$\hat{X}_B - r_B \wedge K_{ca+} \pmod q$$
$$= \left(S_B - \overline{HID_B} \times K_{B+}\right) \times K_{ca+} \pmod q$$

If $H\left(\hat{X}_B\right) = H\left(X_B\right)$ then the node B is verified.

- B computes:

$$\overline{HID_A} = H\left(K_{A+} \| ID_A\right)$$
$$\hat{X}_A - r_A \wedge K_{ca+} \pmod q$$
$$= \left(S_A - \overline{HID_A} \times K_{A+}\right) \times K_{ca+} \pmod q$$

If $H\left(\hat{X}_A\right) = H\left(X_A\right)$ then the node A is verified.

Now the common envelope key is calculated:

$$K_{AB} = f_A^{-1} \times K_{B+} \pmod q$$
$$K_{AB} = f_B^{-1} \times K_{A+} \pmod q$$

B.2 Mutual authentication

The two nodes A, B share the envelope key K_{AB}, and a challenge response protocol is then applied as follows:

- A creates a random polynomial $t_A \in L_k$ and computes

$$T_A = t_A \times K_{ca+} \pmod q$$
$$V_A = K_{AB} + T_A \pmod q$$

Then it sends $\left(ID_A, V_A\right)$ to B.

- B creates a random polynomial $t_B \in L_k$ and computes

$$T_B = t_B \times K_{ca+} \pmod q$$
$$V = K + T_B \pmod q$$
$$\hat{T}_A - T_A \quad K_{AB}$$

Since the identity of B is correct, B will have $\hat{T}_A - T_A$.

To verify this, B will compute

$$W_B = T_B \times \hat{T}_A \pmod q$$
$$I\left(B\right) = H\left(ID_A \| ID_B \| W_B\right)$$
$$Z_B = W_B + K_{AB}$$
$$I\left(A\right)^* = H\left(ID_A \| ID_B \| Z_B\right)$$

And sends $\left(ID_B, V_B, I\left(B\right)\right)$ to A.

- Upon reception of the previous variables, A will authenticate $I\left(B\right)$ by computing

$$\hat{T} - K$$
$$\hat{W}_B - \hat{T}_B \wedge T_A \pmod q$$
$$I\left(B\right)^* = H\left(ID_A \| ID_B \| \hat{W}_B\right)$$

If $I\left(B\right)^* = I\left(B\right)$, then the identity of B is authenticated.

In addition A computes

$$Z_A = \hat{W}_B + K_{AB}$$
$$I\left(A\right) = H\left(ID_A \| ID_B \| Z_A\right)$$

And sends $I\left(A\right)$ to B.

- On receiving $I\left(A\right)$, B will authenticate the identity of A if $I\left(A\right) = I\left(A\right)^*$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Now both A, B are mutually authenticated, and they can compute a common secret key for data encryption simultaneously:

$$SK_A = H\left(ID_A \| ID_B \| T \| \widehat{\ } \ldots \widehat{\ }_B\right)$$

$$SK_B = H\left(ID_A \| ID_B \| \widehat{1}_A \| \ldots_B \| \ldots_B\right)$$

** When authentication fails, A will try to reconnect, in the same way, with another neighbor.

C. Route discovery

The node must ensure that the intended destinations are indeed reached, so each node must maintain a routing table with entries for source-destination active pairs. The route discovery begins with a node broadcasting a route discovery packet (RDP) to its neighbors.

$$A \rightarrow brdcst\left[RDP, IP_X, N_A\right]K_{A-}, Cert_A$$

The RDP includes a packet type identifier "RDP", the IP address of the destination X, A's certificate and a nonce N_A, all signed with A's private key. The purpose of the nonce is to uniquely identify an RDP coming from a source. Each time, A performs route discovery it increases the nonce.

Each node validates the signature with the certificate, updates its routing table with the neighbor from which it received the RDP, signs it, and forwards it to neighbors after removing the certificate and the signature of the previous node.

The initiator's signature and certificate are not removed.

Let B be a neighbor that has received the RDP broadcast from A. It rebroadcasts the RDP with its certificate:

$$B \rightarrow brdcst : \left[\left[RDP, IP_X, N_A\right]K_{A-}\right]K_{B-}, Cert_A, Cert_B$$

Then B's neighbor, which is C, validates the signatures for both the RDP initiator, and B. Then removes B's certificate and signature, signs the contents of the message with its private key, then rebroadcasts the RDP with its certificate.

$$C \rightarrow brdcst : \left[\left[RDP, IP_X, N_A\right]K_{A-}\right]K_{C-}, Cert_A, Cert_C$$

Finally, the message is received by the destination X, who replies to the first RDP received for a source and a given nonce. Because RDP's do not contain a hop count or a specific recorded source route; the discovered route may not be the shortest. However, since the messages are signed at each hop, malicious nodes have no opportunity to redirect traffic.

After receiving the RDP, the destination unicasts a Reply (REP) packet back along the reverse path to the source. Let the first node that receives the REP sent by X be node D.

$$X \rightarrow D : \left[REP, IP_A, N_A\right]K_{X-}, Cert_X$$

The REP contains the address of the source node, the destination's certificate, a nonce, and the associated timestamp. The destination node signs the REP before transmitting it. The REP is forwarded back to the initiating node by a process similar to the process described for the route discovery, except that the REP is unicasted along the reverse path. All data that have been transmitted is not just signed but also encrypted.

Let C be the next hop to the source

$$D \rightarrow C : \left[\left[REP, IP_A, N_A\right]K_{X-}\right]K_{D-}, Cert_X, Cert_D$$

C validates D's signature, removes the signature and certificate of D. Then signs the contents of the message and appends its own certificate before unicasting the REP to B.

$$C \rightarrow B : \left[\left[REP, IP_A, N_A\right]K_{X-}\right]K_{C-}, Cert_X, Cert_C$$

Each node checks the nonce and signature of the previous hop as the REP is returned to the source. When the source receives the REP, it verifies the destination's signature and the nonce returned by the destination.

D. Route Maintanance

The route is de-activated in the route table when no traffic has occurred on it for its specific lifetime. Data received on an inactive route causes nodes to generate an Error message (ERR). Nodes also use ERR messages to report links in active routes that are broken due to node movement. All ERR messages must be signed by related nodes. For a route between source A and destination X, a node B generates the ERR message for its neighbor C as follows:

$$B \rightarrow C : \left[ERR, IP_A, IP_X, N_B\right]K_{B-}, Cert_B$$

This message is forwarded along the path toward the source without modification. A nonce ensures that the ERR message is fresh. It is extremely difficult to detect when ERR messages are fabricated for links that are truly active and not broken. However, the signature on the message prevents impersonation and enables non-repudiation. The node that transmits a large number of ERR messages, whether the ERR messages are valid or fabricated, should be avoided.

V. SECURITY AND EFFICIENCY ANALYSIS

To analyze the efficiency of this protocol; table 1 compares the security levels of NTRU and two well-known cryptosystems: ECC (Elliptic curve cryptography) and RSA (Rivest Shamir, Adleman) [17].

Table 1, security levels of cryptosystems [17]

Security levels	NTRU /MIPS years	ECC /MIPS years	RSA /MIPS years
Moderate security	167×10^6	106×10^4	512×10^4
Standard security	251×10^{12}	160×10^{11}	1024×10^{11}
High security	503×10^{35}	210×10^{20}	2048×10^{20}

Where MIPS-years refers to the time in years we need to break the cryptosystem using a processor with speed measured by "Million Instructions per second (MIPS)".

We can easily notice that NTRU is the best choice, because of its high security and low complexity. The low complexity of NTRU comes from the fact that it uses ring polynomials to encrypt and decrypt which is just addition and multiplication operations,

On the other hand, our protocol efficiently offers a secure fastest path, which is chosen by the RDP that reaches the destination first.

- To analyze the security of the proposed protocol, we show how it behaves against most of the attacks.
  - Brute-force attack

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

If an attacker tries to attack the private key in KG for the node A, it needs to figure out all F *and* f_A, when $F \times K_{ca+} = g \pmod q$ *and* $f_A \times K_{A+} = g \pmod q$ are small polynomials. The attacker has to guess all possible polynomials g, so the complexity of this would be

$$\frac{1}{d_g!}\sqrt{\frac{N!}{(N-2d_g)!}}$$

Which is just complex [16], where g has d_g coefficients equal to 1 and d_g coefficients equal to –1.

- Common session key attack

Attacks to common key K_{AB} need to figure out all

f_A *and* f_B

$f_A \times K_{A+} = g \pmod q, f_B \times K_{B+} = g \pmod q$

As the previous attack, the complexity would be the same as before

$\frac{1}{d_g!}\sqrt{\frac{N!}{(N-2d_g)!}}$ and the attack will fail.

- Man in the middle attack

Attacker might hack (ID_A, X_A, S_A, K_{A+}) and forge the identity of A. However, this attacker will not be able to pass the authentication phase since ID_A is protected under the hash function which prevents the attacker from computing $\hat{X}_A - X_A$ correctly. Furthermore, it will be computationally intensive for an attacker to figure out k_A. In NTRU cryptosystem the combination of L_f, L_g, L_k will be of the order of 2^{160} even for low key security(moderate) $(N, p, q) = (167, 3, 128)$, $key\ security \approx 2^{80}$. Hence, it is secure against this attack

- Attacks using modification

Our proposed protocol specifies that all fields of RDP and REP packets remain unchanged between source and destination. Since the initiating node signs both packet types, any alterations during transit would be detected, and the altered packet would be subsequently discarded. Thus, modification attacks are prevented.

For attacks with modified hop-count, the packets in this protocol do not contain hop-count field, so it is secured against this attack too.

- Attacks using Impersonation

Route discovery packets contain the certificate of the source node and are signed with the source's private key. In addition, reply packets include the destination node's certificate and signature, so our protocol prevents impersonation attacks where either the source or the destination node is spoofed.

- Attacks using Fabrication

All the routing messages in this method must include the sending node's certificate and signature, so it ensures non-repudiation and prevents spoofing and unauthorized participation in routing. A node that continues to inject false messages into the network may be excluded from future route computation.

Obviously, the proposed technique is much more effective than the traditional ones, mentioned in paragraph (II-B).The complexity is simple because of the use of NTRU, which uses only additions and multiplications. It is also safer, because the attacker will not be able to afford the time spent to get any information. The most important point in this protocol is that it does not need a third party to authenticate any node in the network, or to distribute keys, so we can reach high levels of operational confidentiality.

VI. RESULTS AND DISCUSSION

A. AVISPA

"Automated Validation of Internet Security Protocols" is a project developed to analyze and test security of protocols by [19]:

- The research team led by A. Armando at the University of Genova, Italy.
- The team led by M. Rusinowitch at INRIA-Lorraine, Nancy, France.
- The team led by D. Basin at the ETH Zurich, Switzerland.
- The team led by J. Cuellar at SIEMENS AG, Munich, Germany.

The structure of the AVISPA tool is described in

Figure 1. The language we use is HLPSL "High Level Protocol Specification Language" which defines the global variables, information types and security values. Once the code is written, the intermediate format (IF) module compiles the code and builds the scenario for testing.

Then, the output of (IF) goes to one of four applications to demonstrate the scenario.

- On-the-Fly-Model-Checker OFMC:
  This application will copy the protocol and run it with the defined parameters including the parameters of nodes to decide if there is any weak point to attack.
- Constraint-Logic-based Attack Searcher CL-AtSe:
  This application will apply the constraints we put on the protocol to decide if the protocol is safe against the attack.
- SAT-based Model Checker SATMC:
  In this application, the security constraints will be added to the protocol and tested against any type of unknown model (as an attacker).
- Tree automata for analysis; the Security Protocol TA4SP protocol analyzer:
  Here, the application will estimate the security of the protocol using Dolev-Yao model [20].

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

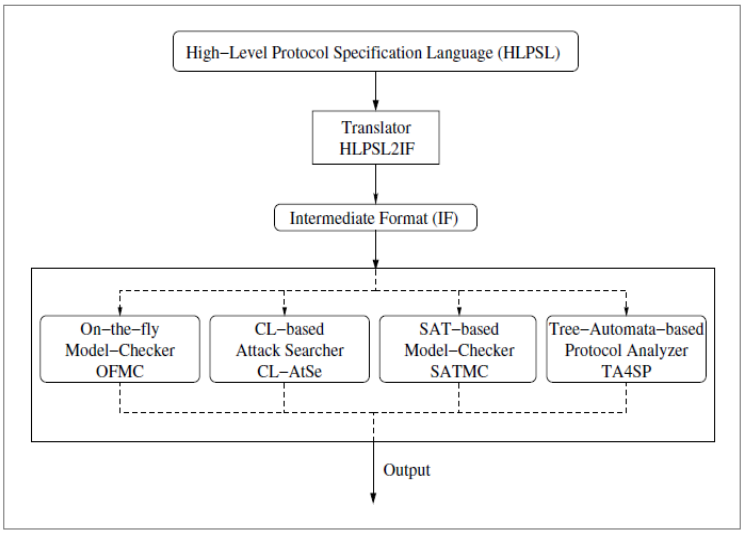


Figure 1 AVISPA structure

After implementing the four applications, we have to represent and simulate the protocol, identify possible attack types against it, and assess, whether or not, it is able to prevent those attacks. Thus, we can conclude whether the protocol is safe under pre-defined conditions.

To test security for any protocol, HLPSL defines events to describe nodes' authentication behaviors, mainly; there are witness and wrequest, which we use in weak authentication. Weak authentication is defined as authentication between two nodes without need to third party.

- Witness (A,B,v,M): A is a witness to authenticate B by checking M value using v protocol.
- Wrequest (B,A,v,M): B requests to be authenticated by M value using v protocol, since A has already been a witness for that value before.

B. Simulation and Results

The proposed traditional scenario simulation is given in Figure 2. Where there is one transmitter node A, some intermediate nodes B, a receiver node X and the intruder node I.

The intruder node I will pretend being an intermediate node, so the messages in the network would be:

$$A \rightarrow I : \{RDP, X, N_A\}_{K_{A-}}, cert_A$$

$$I \rightarrow B : \{RDP, X, N_A\}_{K_{A-}}, cert_A$$

$$B \rightarrow I : \{\{RDP, X, N_A\}_{K_{A-}}\}_{K_{B-}}, cert_A, cert_B$$

When B moves out of the network coverage area, node I would try to locate between A and X, as an intermediate node to the X.

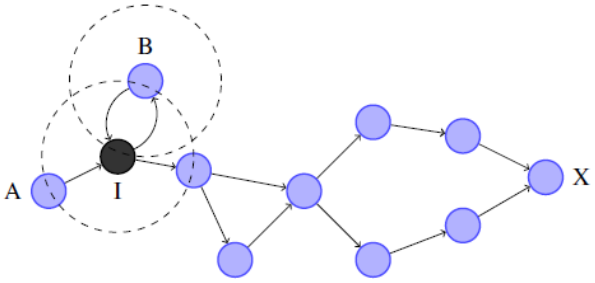


Figure 2 traditional scenario for simulation

The role is defined as a module with basic information and transactions between other nodes. Roles for all nodes in the network can be divided into 3 main roles:

- The first role is the **"source"**, which has two tasks. One of them is to broadcast RDP messages, and the other is to receive REP messages from the destination directly or through intermediate nodes.

The initial state of the source is defined as follows:

- A is the source, B and C are intermediate nodes and X is the destination.
- K_a is the public key of A and K_x is the public key for X.
- Channel parameters are SND and RCV[4].

In next state, the source broadcasts RDP message with special number N_a (using its key), and waits for a response.

From security point of view, in the "transmit" mode, A has to be witness for the address of destination X and the special number N_a. but in the "receive" mode, A would request security approve for its address and the special number N_a. see Figure 3

```
role source_node(
    A,B,C,X : agent,
    Ka, Kb, Kc, Kx : public_key,
    RCV, SND : channel(dy))
played_by A def=
    local
        RDP,REP : protocol_id,
        State   : nat,%{not assigned}
        NA      : text %identifier
    const
        a_b_IPa, a_b_IPx, a_a_Na : protocol_id
    init State := 0
    transition
        step1.
          State = 0 /\ RCV(start)  %state=0  and starting command
          =|>
          State':= 4 /\ Na' := new() % updated values needs
          /\ SND({RDP.X.Na'}_inv(Ka))  /\ witness(A,B,a_b_IPx,X)
          /\ witness(A,B,a_a_Na,Na')
        step2.
          State = 4 /\ RCV({{REP.A'.Na'}_inv(Kx)}_inv(Kb))
          =|>
          State':= 8 /\ wrequest(A,B,a_b_IPa,A') /\ wrequest(A,B,a_a_Na,Na')
end role
```

Figure 3 the source node description using HLPSL

- The second role is the **"destination"**, which has just one task. It has to reply to RDP messages, which can arrive directly from the source, or through the intermediate nodes. The destination node has to be witness for the address of source node A and the special number N_a; also, it has to be request for its address and the special number N_a, see Figure 4

- The third role is **"intermediate"**. There are two types of intermediate nodes. The first one contains the source's neighbors, and the second type contains the far away nodes from the source A. In the first type, the node B must verify the A's certificate, also it has to be witness for (X,N_a) to the next node and request for the previous node, which is A.. In the second type, the node B must verify two certificates; one for the source A, and the other for the neighbor node C.

[4] Send and receive

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Also, node B must be witness for the next node C and request for the previous one.

```
role destination_node(
      A,B,C,X : agent,
      Memory : (text.text.agent) set,
      Ka, Kb, Kc, Kx : public_key,
      RCV,SND : channel(dy))
played_by X def=
      local
        RDP,REP : protocol_id,
        State : nat,
        Na : text
      const
         a_b_IPa, a_b_IPx, a_a_Na : protocol_id
      init State := 3
      transition
          step1.
            State = 3 /\ RCV({{RDP.X'.Na'}_inv(Ka)}_inv(Kc)
                     /\ not(in(RDP.Na'.A', Memory))
            =|>
            State':= 7 /\ SND({REP.A'.Na'}_inv(Kx)
                        /\ wrequest(X,C,a_b_IPx,X') /\ wrequest(X,C,a_a_Na,Na')
                        /\ witness(X,C,a_b_IPa,A') /\ witness(X,C,a_a_Na,Na')
end role
```

Figure 4 the destination node description using HLPSL

Another issue for intermediate node is that it must firstly verify the duality (X,N_a) is not in its memory; otherwise, the message will be discarded as it had passed before, see Figure 5

- There are two more roles to complete this simulation; session role and environment role. In the session role, we define the main initial topology of the network, which consists of one source A, one destination X and two intermediate nodes B,C, as it is shown below in Figure 6 and Figure 7.

```
role intermediate_node2(
    A,B,C,X : agent,
    Memory : (text.text.agent) set,
    Ka, Kb, Kc, Kx : public_key,
    RCV,SND : channel(dy))
played_by C def=
    local
        RDP,REP : protocol_id,
        State : nat,
        Na : text
    const
        a_b_IPa, a_b_IPx, a_a_Na : protocol_id
    init State := 2
    transition
        step1.
            State = 2 /\ RCV({{RDP.X'.Na'}_inv(Ka)}_inv(Kb)
                     /\ not(in(RDP.Na'.X', Memory))
            =|>
            State' := 6 /\ SND({{RDP.X'.Na'}_inv(Ka)}_inv(Kc). % two
            verification processes
                        /\ wrequest(C,B,a_b_IPx,X') /\ wrequest(C,B,a_a_Na,Na')
                        /\ witness(C,X,a_b_IPx,X') /\ witness(C,X,a_a_Na,Na')
        step2.
            State = 6 /\ RCV({REP.A'.Na'}_inv(Kx)) /\ not(in(RDP.Na'.X',
            Memory))
            =|>
            State':= 10 /\ SND({{REP.A'.Na'}_inv(Kx)}_inv(Kc).
                        /\ wrequest(C,X,a_b_IPa,A') /\ wrequest(C,X,a_a_Na,Na')
                        /\ witness(C,B,a_b_IPa,A') /\ witness(C,B,a_a_Na,Na')
end role
```

Figure 5 the intermediate node description using HLPSL

```
role session(
   A,B,C,X : agent,
   Ka, Kb, Kc, Kx : public_key)
   def=
   local
      Mem1,Mem2,Mem3 : (text.text.agent) set,
      RCV1,SND1,RCV2,SND2,RCV3,SND3,RCV4,SND4 : channel(dy)
      Init Mem1:= {} /\ Mem2:= {} /\ Mem3:= {}
   composition
      source_node(A,B,C,X,Ka,Kb,Kc,Kx,RCV1,SND1)
      /\ intermediate_node1(A,B,C,X,Mem1,Ka,Kb,Kc,Kx,RCV2,SND2)
      /\ intermediate_node2(A,B,C,X,Mem2,Ka,Kb,Kc,Kx,RCV3,SND3)
      /\ destination_node(A,B,C,X,Mem3,Ka,Kb,Kc,Kx,RCV4,SND4)
end role
```

Figure 6 the session role description using HLPSL

In the environment role, we have to define the whole topology of the network using session role. We also define the intruder's knowledge in this network. This knowledge depends on the NTRU algorithm.
When defining the topology with multiple options, we have to test them all to decide whether the proposed protocol is safe or not. As shown in Figure 7 below, there are eight healthy nodes and one mal node, which will take the place of one of the healthy nodes.

```
role environment()
def=
const
      a_b_IPa, a_b_IPx, a_a_Na : protocol_id,
      a,b,c,x : agent,
      ka,kb,kc,kt,kx : public_key
      intruder_knowledge ={a,b,c,x,ka,kb,kc,kx}
composition
      session(a,b,c,x,ka,kb,kc,kx)
      /\ session(a,i,c,x,ka,kb,kc,kx)
end  role
```

Figure 7 the environment role description using HLPSL

- Now by compiling the previous roles and pass them to the AVISPA applications we can see that the protocol is safe for each topology. In Figure 8, we can see the discovering path model for the network from node a-3 to node x-11 and in Figure 9 the intruder has joined to the network.

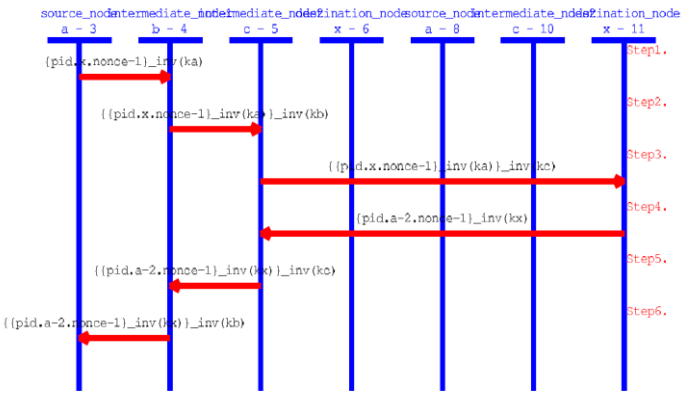


Figure 8 discovering path from a-3 to x-11

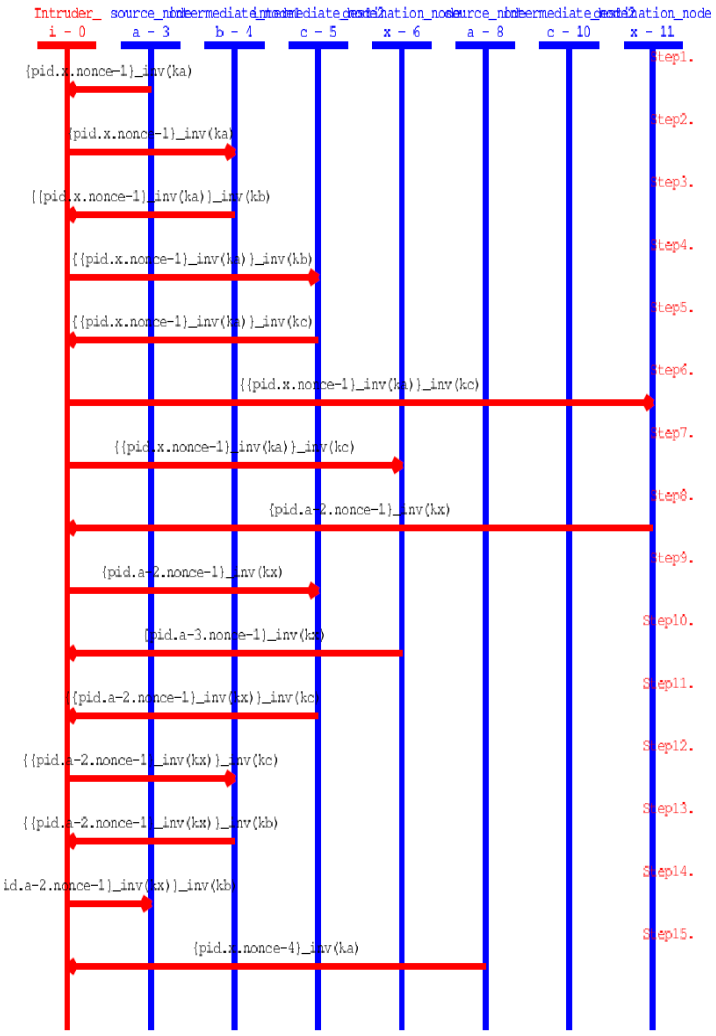


Figure 9 the attack model simulation

- After simulating this model, the OFMC application will tell us that the proposed protocol is safe under the security conditions we defined and cannot find a weak point to attack. This is shown in Figure 10

```
% OFMC
% Version of 2006/02/13
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
PROTOCOL
  /home/span/span/testsuite/results/proposed.if
GOAL
  as_specified
BACKEND
  OFMC
COMMENTS
STATISTICS
  parseTime: 0.00s
  searchTime: 0.53s
  visitedNodes: 68 nodes
  depth: 9 plies
```

Figure 10 the result of OFMC application

- Moreover, the CL-AtSe will tell us that the proposed protocol is safe against that intruder. See Figure 11

```
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
  TYPED_MODEL
PROTOCOL
  /home/span/span/testsuite/results/proposed.if
GOAL
  As Specified
BACKEND
  CL-AtSe
STATISTICS
    Analysed   : 49 states
  Reachable  : 17 states
  Translation: 0.06 seconds
  Computation: 0.00 seconds
```

Figure 11 the result of CL-AtSe application

VII. Conclusion and Further Work

This paper has presented an authenticated routing protocol for wireless ad-hoc networks. A traditional authenticated routing protocol requires a trusted third party for obtaining certificates; and thus it is not preferable for applications where there is no existing infrastructure like MANET.

In view of securing on demand routing for MANET, we have proposed an authenticated routing protocol depending on NTRU cryptography. This technique is much more efficient because it only requires computation of additions and multiplications. In addition, it is secure enough because the attacker needs non-affordable time to complete the attack in polynomials ring. The proposed protocol has advantages over protocols with similar objectives because of the fast and extremely secure user identification system. The model does not need an on-line certificate authority for key distribution. The security proofs of the protocol are done through use of AVISPA tool and the results are shown in the paper.

On the other hand, as the proposed model concerns on-demand protocols, it has the benefits of high performance and low cost due to its eventual reactive nature. In addition, our proposed model provides solutions for some attacks but it is silent about some other attacks -like black hole attack-. Future research can be conducted to add other functionalities to it.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

References

[1]. *Du, Xinjun, Ge. Jianhua, Wang, Ying, "A method for security enhancements in AODV Protocol", In Proceedings of the17th International Conference on Advanced Information Networking and Applications (AINA'03), IEEE Computer Society Washington, DC,* ***2003***.

[2]. *B.Karthikeyan, N.Kanimozhi, S. Hari Ganesh, "Analysis of Reactive AODV Routing Protocol for MANET", 2014 World Congress on Computing and Communication Technologies, IEEE,* ***2014***.

[3]. *S. Capkun, L. Buttyan, J.-P. Hubaux. "Self-Organized Public-Key Management for Mobile Ad Hoc Networks". Technical Report EPFL/IC/200234, Swiss Federal Institute of Technology, Lausanne, June* ***2002***.

[4]. *Sh. Rani1, P. Singh and R. Preet. "Reviewing MANETs & Configurations of Certification Authority (CA) for node Authentication". International Journal of Computer Science and Information Technologies, Vol. 4 (6),* ***2013***.

[5]. *H. Vegda, N. Modi "Review Paper on Mobile Ad-hoc Networks". International Journal of Computer Applications, Volume 179 – No.37, April* ***2018***.

[6]. *Ashish Sharma, Dinesh Bhuriya, Upendra Singh, "Secure Data Transmission on MANET by Hybrid Cryptography Technique", IEEE ,* ***2015***.

[7]. *Utpal Kumar Verma, Sushil Kumar, Ditipriya Sinha, "A Secure and Efficient Certificate based Authentication Protocol for Manet", IEEE,* ***2016***.

[8]. *F. Jasdanwala, N. Dutta, D. Bhattacharyya "Lattice Based Secure Data Transmission in MANETs". IEEE International Conference on Computational Intelligence and Computing Research,* ***2014***.

[9]. *S. Tan, P. Sok, K. Kim "Using Cryptographic Technique for Securing Route Discovery and Data Transmission from BlackHole Attack on AODV-based MANET". International Journal of Networked and Distributed Computing, Vol. 2, April* ***2014***

[10]. *RB. Sajyth, G. Sujatha "Design of data confidential and reliable bee clustering routing protocol in MANET" International Journal of Engineering & Technology, 7 (2.8),* ***2018***.

[11]. *V. Kaur, S. Rani "A Hybrid and Secure Clustering Technique for Isolation of Blackhole Attack in MANET". International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 7, Issue 3, March* ***2018***.

[12]. *U. Verma, S. Kumar, D. Sinha "A Secure and Efficient Certificate based Authentication Protocol for Manet". International Conference on Circuit, Power and Computing Technologies [ICCPCT],* ***2016***.

[13]. *R. Shaktawat, D. Singh, N. Choudhary, "An Efficient Secure Routing Protocol in MANET Security - Enhanced AODV (SE-AODV)". International Journal of Computer Applications. Volume 97– No.8, July* ***2014***.

[14]. *M. Kumar, M. Faisal, A. Ahmed, "Attacks in MANET". International Journal of Research in Engineering and Technology. Volume 02- Issue 10, Oct* ***2013***.

[15]. *Abd. Ali and U.V. Kulkarni. "A State-of-The-Art of Routing Protocols for Mobile AdHoc Networks (MANET)". International Journal of Computer Applications, Vol.127,* ***2015***.

[16]. *S. Aluvala, K. Raja Sekhar and D. Vodnala. "An Empirical Study of Routing Attacks in Mobile Ad-hoc Networks". 2nd International Conference on Intelligent Computing, Communication & Convergence (ICCC),* ***2016***.

[17]. *E. Huei Lu, HK. Chang, SH Liaw "A Security and Efficiency of Authenticated Key Exchange Protocol for Wireless Mobile Ad Hoc Networks". Trans Tech Publications, Switzerland, Applied Mechanics and Materials Vols. 284-287,* ***2013***.

[18]. *E. Okamoto, K. Tanaka, "Identity-based Information Security Management System for Personal Computer Networks," IEEE J. Select. Areas Commun. vol. 7, Feb.* ***1989***.

[19]. *The AVISPA Project. URL: www.avispa-project.org.*

[20]. *D. Dolev and A. Yao. "On the Security of Public-Key Protocols". IEEE Transactions on Information Theory, 2(29),* ***1983***.

AUTHORS PROFILES …

Alaa MOUALLA, Engineering Degree in Telecommunication Systems, from the Higher Institute of Applied Sciences and Technology (HIAST), Damascus, Syria, 2007. Master Degree in Computer Networks from the Higher Institute of Applied Sciences and Technology (HIAST), Damascus, Syria, 2016. A PhD student in Communication Network Security in HIAST.
He has interests in Cryptography and Network Security and teaches these subjects in "Communication Systems" master in HIAST as an assistant teacher, in addition to his interests in communication, Information and Data Processing.
He is a Lecturer at the Higher Institute of Applied Sciences and Technology (HIAST) and Damascus University.

Oumayma ALDAKKAK, Engineering Degree in Systems' Electronics, National High School of Electronics and Radio-Electricity in Grenoble (ENSERG), France, 1985. Postgraduate Degree (DEA) in Electronic Systems (ENSERG), France, 1985 also. A PhD in Electronic Systems from "Speech Communication Institute/Institut de la Communication Parlée"-"Institut National Polytechniques de, Grenoble" (ICP-INPG), France, 1988.
She has interests in Cryptography and Data Security and teaches theses subjects in "Communication Sytems" master in HIAST, in addition to her interests in Speech and Natural Language Processing domain, and Arabic Language Resources.
Prof. Al Dakkak is a Research Director at Higher Institute for Applied Sciences and Technology HIAST, Lecturer at the Information Technology Engineering Faculty, Damascus University. And in the Syrian Virtual University (SVU).. Former Head of Communication Department in HIAST (2007-2014).

Mohamad ALJNIDI, Engineering Degree in Compyter Science from the Higher Institute of Applied Sciences and Technology (HIAST), Damascus, Syria, 1996. Master Degree in Computer Networks from Pierre et Marie Curie University (Paris VI), Paris, France, 2005. And a PhD in Computer Network Security from TELECOM ParisTech (ENST: Ecole Nationale Superieurs des Telecommunications), Paris, France, 2009.
He has interests in Information and Network Security and teaches theses subjects in "Communication Sytems" Master in the Higher Insttute of Applied Sciences and Technology (HIAST), in addition to his interests in Autonomic Communications, Software Defined Networks and Cloud Computing.
He is an assistant professor at the Higher Insttute of Applied Sciences and Technology (HIAST), the Arab Academy for E-Business (ARAEB), and the Syrian Virtual University (SVU). He is a Member of the administrative committee in the Syrian Computer Society (SCS).

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Detection of Primary User at Fusion Center of a CRN Using Fuzzy-Logic Rules

Md Abul Kalam Azad
Professor, Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh, e-mail: makazad@juniv.edu

Md. Imdadul Islam
Professor, Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh, e-mail: imdad@juniv.edu

Sanjit Kumar Saha
Assistant Professor, Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh, e-mail: sanjit@juniv.edu

Jugal Krishna Das
Professor, Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh, e-mail: cedas@juniv.edu

***Abstract*— The performance of a CRN (cognitive radio network) solely depends of detection of a primary user (PU) under spectrum sensing on the traffic channel. In a fusion center (FC) of a CRN the SNR (signal to noise ratio) of several CR are combined under combining scheme (MRC, EGC or SC) to take the decision of presence or absence (called Hypothesis H_1 or H_2) of a PU on a particular traffic channel. In this paper we apply Fuzzy-logic rules on received SNR of each CR at the fusion center under awgn (additive white Gaussian noise). Like previous work, three Gaussian membership functions are used for SNR (low, moderate and high) of each CR then 50 Fuzzy rules are applied on the relative magnitudes SNR to decide the situation of hypothesis H_0 or H_1. In this paper we extend the job using the combination of membership functions of SNR, SIR (signal to interference ratio) and SINR (signal to noise plus interference ratio) which improves the detection hypothesis at the expense of process time and complexity of Fuzzy model.**

Keywords- Co-operative CRN, SIR, SNR, surface plot and fuzzy rules.

I. INTRODUCTION

Fuzzy system has versatile applications but this paper only concentrated about its applications in co-operative CRN. In [1] authors uses Fuzzy Logic system in spectrum access of CRN. Three inputs are used in fuzzy system, they are 'Spectrum utilization efficiency', 'Degree of mobility', 'Distance from primary user to the secondary users' and each input has three levels: low, moderate, and high as the membership function. The output of the fuzzy system /consequence has five levels and call arrival rate is related with different parameters of CRN. Similar concept is available in [2]; where spectrum handoff algorithm (the SU returns the channel to a PU when it claims the channel) is implemented using fuzzy system. More elaborate explanation of spectrum handoff is available in [3] where three MF is used: PU interference, SINR and Handoff status is used to take the decision of spectrum handoff. In [4] authors deals with Spectrum Availability Assessment of CRN where four inputs are used in the Fuzzy decision-making system: Detection probability, operational SNR, available time and priori information. The first and fourth membership function has three levels: low, medium and high but the second and third MF has two levels: low and high. Then applying 36 rules the de-fuzzification provides: Energy detection, correlation detection, Feature detection, match filtering, co-operative energy detection, channel change etc. Finally the system output takes decision spectrum sensing technique. The extension of the work is found in [5] where both average and total throughput (in Mbps) of SU are determined. Here four MFs: spectrum utilization efficiency, mobility, distance and signal strength are used as input variable and priority factor is used as the output of the fuzzy system. Each of the four input MFs has three levels: Low, Medium, High and the output has levels: Very Low, Low, Medium and High. Total 81 rules are applied to get output then compared with previous work of 'Kaniezhil's scheme' and found better result. Similar concept is also available in [6-8] for co-operative CRN.

In [8] five layers ANFIS model is used in spectrum sharing of CRN where the two input of the system are: SNR of PU and PU's interference channel gain. The authors use trapezoidal membership functions for both inputs and each MF has three levels. The performance of the SU is determined using from the profile of SER vs. SNR for both with and without ANFIS. The performance is fount better with ANFIS model at moderate and high SNR. In this paper we use Gaussian membership function on SNR, SIR and SINR based integrated spectrum sensing model with more than 200 rules. The Fuzzy model of the paper can be used in FC to detect the presence of a PU under adverse condition of the wireless link.

The rest of the paper is organized as: section 2 provides the basic statistical model of signal analysis of both CRN and co-operative CRN, section 3 proves fuzzy system of co-operative CRN based on SNR, SIR and SINR; section 4 provides results based on analysis of fuzzy system of section 3 and section 5 concludes entire analysis.

II. SIGNAL STATISTICS OF CRN

In CRN the primary or licensed user (PU) can access the spectrum of traffic channel when it is not used by other PU. The SU users are opportunistic users have lower priority and

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

can access a traffic channel only when it is remains unused by PUs. The main objective of a PU is to search for free channels using spectrum sensing algorithm even when PU claims a channel, the SU immediately releases it.

In a CRN the received signal of a user under awgn is modeled using two hypothesis like [10-11],

$$r(t)=\begin{cases} n(t) & \text{Under } H_0 \\ s(t)+n(t) & \text{Under } H_1 \end{cases} \quad (1)$$

;Where $s(t)$ is the transmitted signal and $n(t)$ is the awgn. The receiver circuit is simply a demodulator as shown in Fig.1; where the output of the first correlator is,

$$r_0=\int_0^{T_b} r(t)s(t)dt=\int_0^{T_b}\{s(t)+n(t)\}s(t)dt$$
$$=\int_0^{T_b} s^2(t)dt+\int_0^{T_b} s(t)n(t)dt=E+n \quad (2)$$

;where T_b is the duration of a symbol, E is its energy and n is noise. The output of the second correlator is,

$$r_1=\int_0^{T_b} r(t)s(t)dt=\int_0^{T_b}\{0+n(t)\}s(t)dt$$
$$=0+\int_0^{T_b} s(t)n(t)dt=0+n=n \quad (3)$$

Assuming signal and noise are uncorrelated, we can put $n\approx 0$ and the input signal at the detector is,

$$r=\begin{cases} n & \text{under } H_0 \\ E+n & \text{under } H_1 \end{cases} \quad (4)$$

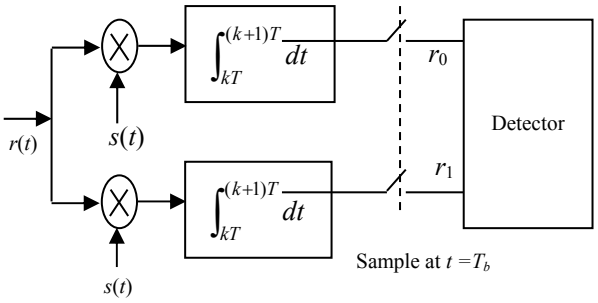


Figure 1. The cross-correlator of received signal

The conditional pdf of received signal under two hypothesis in an awgn channel are,

$p(r|H_0)=\frac{1}{\sqrt{2\pi}\sigma}e^{-r^2/2\sigma^2}$ No PU on traffic channel

and $p(r|H_1)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(r-E)^2/2\sigma^2}$ Presence of PU

shown in Fig. 2.

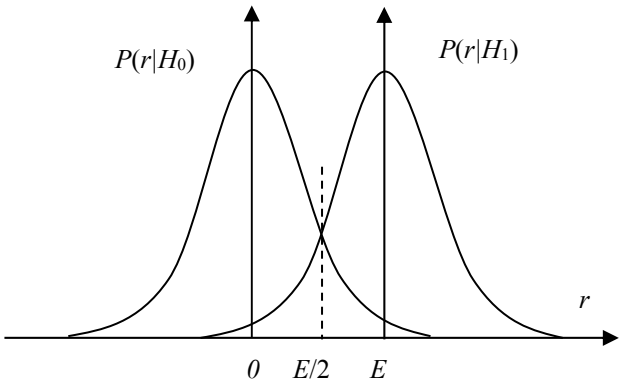


Figure 2. The pdf of $P(r|H_0)$ and $P(r|H_1)$

The probability of error under H_1 is,

$$p_{e1}=P(r<E/2)=\int_{-\infty}^{E/2}\frac{1}{\sqrt{2\pi}\sigma}e^{-(r-E)^2/2\sigma^2}dr \quad (5)$$

The probability of error under H_0 is,

$$p_{e0}=P(r>E/2)=\int_{E/2}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-r^2/2\sigma^2}dr \quad (6)$$

In cooperative spectrum sensing technique of CRN, several SUs conveys the signal of PU to an FC provided each SU detects the spectrum of PU independently like Fig. 3. The fusion center works as a combiner, using fusion rule such as: MRC, SC, AND Logic, OR Logic can detect the presence of PU found in [12-13].

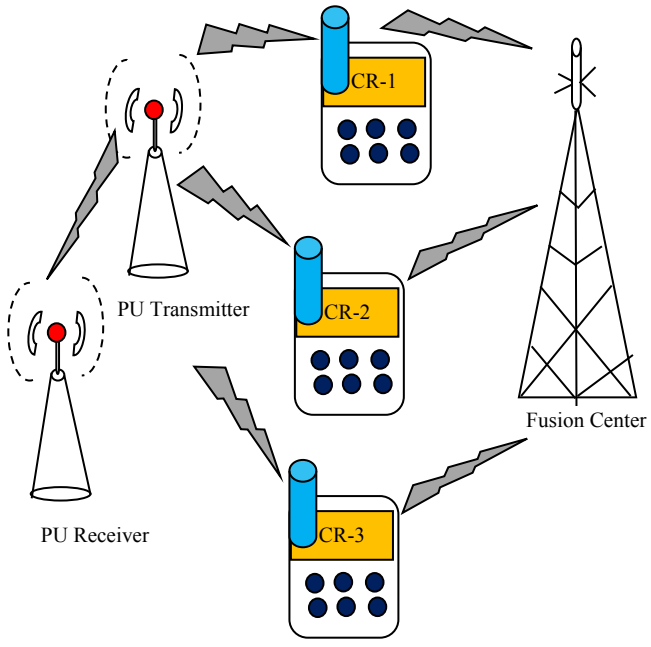


Figure 3. Co-operative spectrum sensing of CRN

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

If the FC takes the decision using AND logic then the probability of detection considering N CRs is,

$$P_{D\_FC} = P_D^{CR-1} \cap P_D^{CR-2} \cap P_D^{CR-3} \ldots \ldots \ldots \cap P_D^{CR-N} = P_D^N \quad (7)$$

Similarly for OR logic case,

$$\begin{aligned} P_{D\_FC} &= P_D^{CR-1} \cup P_D^{CR-2} \cup P_D^{CR-3} \ldots \ldots \ldots \cup P_D^{CR-N} \\ &= \left\{\left(1-P_D^{CR-1}\right) \cap \left(1-P_D^{CR-2}\right) \cap \left(1-P_D^{CR-3}\right) \ldots \ldots \ldots \cap \left(1-P_D^{CR-N}\right)\right\}^c \\ &= 1-\left(1-P_D\right)^N \end{aligned} \quad (8)$$

and for majority decision case we can use the concept of binomial pdf like,

$$P_{D\_FC} = \sum_{j=\left\lceil \frac{N}{2} \right\rceil}^{N} \binom{N}{j} P_D^j \left(1-P_D\right)^{N-j} \quad (9)$$

Above relations can be implemented using fuzzy rule will be shown in next section.

III. Fuzzy System of CRN

While describing human reasoning, logic based on two truth values "true" and "false" is not adequate in most of the cases. "Is X honest?" – The answer to this query need not be definitely "true" or "false". Considering the degree to which one knows X, a variety of answers spanning a range, such as "extremely honest", "extremely dishonest", "very honest" could be generated. The situation is therefore so fluid that it can accept values between 0 and 1, in contrast to the logic bases, which was either "true" or "false". Such a situation is termed fuzzy. Fuzzy logic uses the whole interval between 0 and 1 for describing human reasoning. It is determined as a set of mathematical principles for knowledge representation based on degree of membership and degree of truth. Such a set is termed as fuzzy set. To represent the fuzzy set in a computer, it can be expressed as a function and then mapped the elements of the set to their degree of membership. This function is called membership function and the membership function values need not be described by discrete values always. Sometimes, this turns out to be described by a continuous function. Different shapes of membership functions exists and the shapes could be triangular, trapezoidal, curved and their variations [14]. If X is a universe of discourse and x is a particular element of X, then a fuzzy set A can be defined a collection of ordered pairs $A = \{(x, \mu_A(x)), x \varepsilon X\}$, where $\mu_A(x)$ is a membership function which is associated with the fuzzy set A such that the function maps every element of the universe of discourse X to the interval [0, 1].

In this paper we model a fusion center of co-operative CRN where the level of SNR of several CRs are combined to take the decision of hypothesis H_0 or H_1 shown in Fig. 4 for three CRs case based on the concept of [15-16]. The MF of each CR consists of three parts: Low (L), Moderate (M) and High (H) follows the profile of Gaussian shape since the pdf of received signal follows Gaussian pdf under awgn.

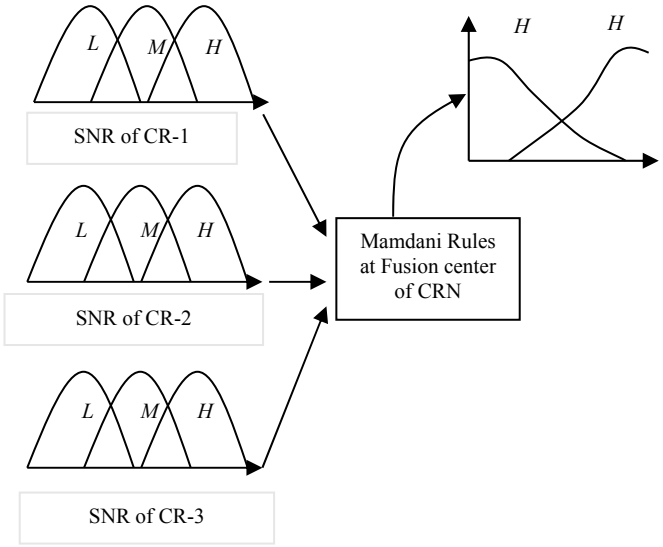


Figure 4. Fuzzy model of a FC of a CRN

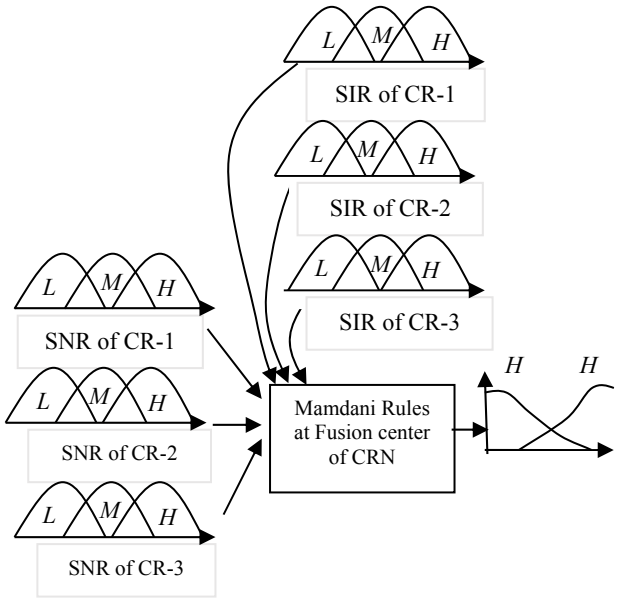


Figure 5. Fuzzy model of a FC of a with SNR and SIR

The second part of the paper deals with combination of SNR and SIR; where the first three CRs are used to detect the presence of PU based on SNR and the rest three CRs based on SIR at the FC shown in Fig. 5. In this case the fuzzy system becomes more complicated and involves more fuzzy rules. The third part of the paper deals with fuzzy system with combination of SNR, SINR and SIR like Fig.6.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

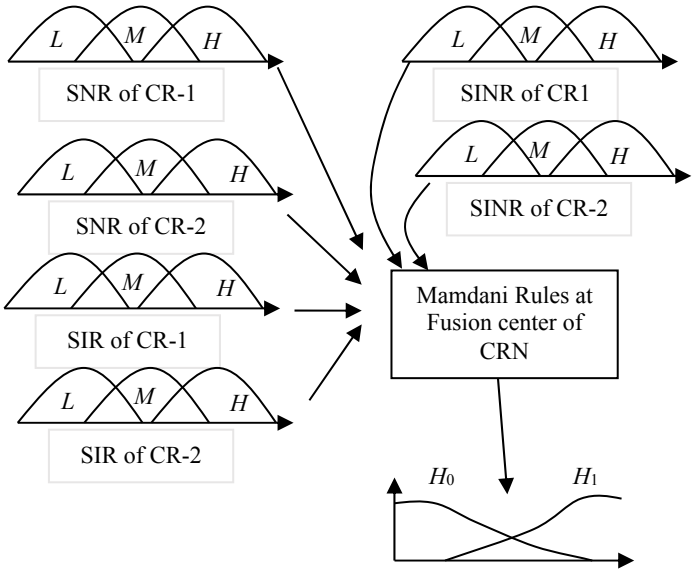


Figure 6. Fuzzy model of a FC of a with SNR, SINR and SIR

IV. RESULTS

First of all we take a simple numerical example where SNR of two CRs are taken as input variable. Here three membership functions of low, moderate and high SNR are used for each of CR. The output deals with two hypotheses H_0 and H_1 pertinent to the case of absence or presence of PU. Next we apply Mamdani rules of 'and' and 'or' logic like:

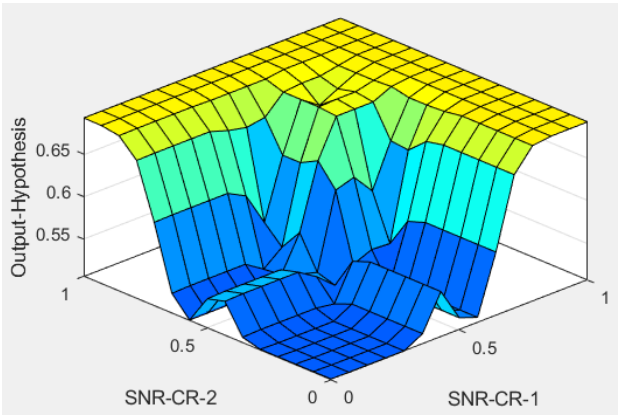
Rule *i*: if u is x_i **and** v is y_i then w is s_i; i = 1, 2, 3,… … …,n;
where $u \in U$, $v \in V$ and $w \in W$.

Rule *j*: if u is x_j **or** v is y_j then w is s_j; j= 1, 2, 3,… … …,m;
where $u \in U$, $v \in V$ and $w \in W$.

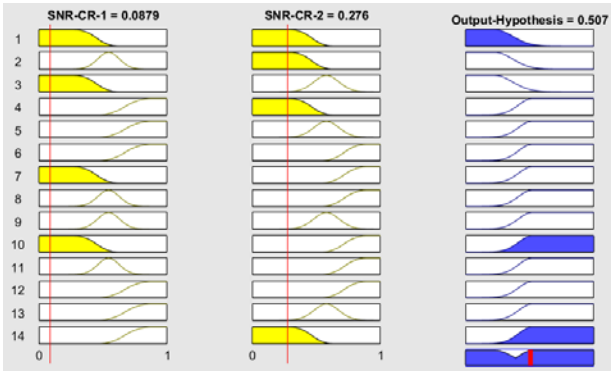
Fig. 7(a) shows 14 rules, (b) shows the surface plot (c) fuzzy rules with SNR-1 is low and SNR-2 is high (d) fuzzy rules when both SNR are moderate (e) fuzzy rules with SNR-1 is moderate and SNR-2 is high (f) both SNR are high.

1. If (SNR-CR-1 is L) and (SNR-CR-2 is L) then (Output-Hypothesis is H0) (1)
2. If (SNR-CR-1 is M) and (SNR-CR-2 is L) then (Output-Hypothesis is H0) (1)
3. If (SNR-CR-1 is L) and (SNR-CR-2 is M) then (Output-Hypothesis is H0) (1)
4. If (SNR-CR-1 is H) and (SNR-CR-2 is L) then (Output-Hypothesis is H1) (1)
5. If (SNR-CR-1 is H) and (SNR-CR-2 is M) then (Output-Hypothesis is H1) (1)
6. If (SNR-CR-1 is H) and (SNR-CR-2 is H) then (Output-Hypothesis is H1) (1)
7. If (SNR-CR-1 is L) and (SNR-CR-2 is H) then (Output-Hypothesis is H1) (1)
8. If (SNR-CR-1 is M) and (SNR-CR-2 is H) then (Output-Hypothesis is H1) (1)
9. If (SNR-CR-1 is M) and (SNR-CR-2 is M) then (Output-Hypothesis is H1) (1)
10. If (SNR-CR-1 is L) or (SNR-CR-2 is H) then (Output-Hypothesis is H1) (1)
11. If (SNR-CR-1 is M) or (SNR-CR-2 is H) then (Output-Hypothesis is H1) (1)
12. If (SNR-CR-1 is H) or (SNR-CR-2 is H) then (Output-Hypothesis is H1) (1)
13. If (SNR-CR-1 is H) or (SNR-CR-2 is M) then (Output-Hypothesis is H1) (1)
14. If (SNR-CR-1 is H) or (SNR-CR-2 is L) then (Output-Hypothesis is H1) (1)

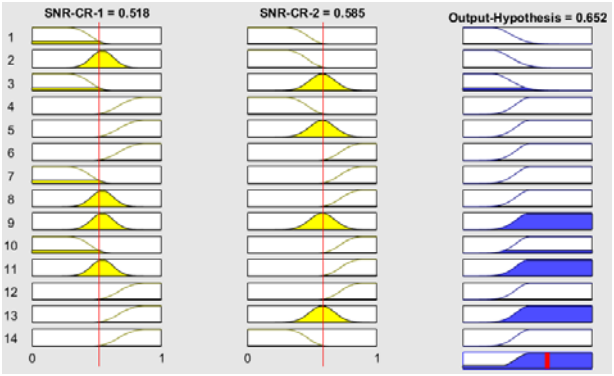
(a)



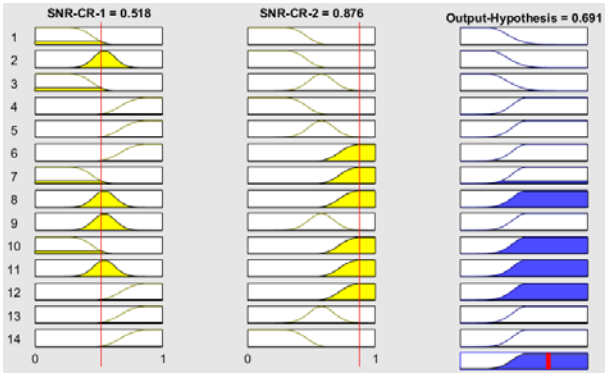
(b)



(c)

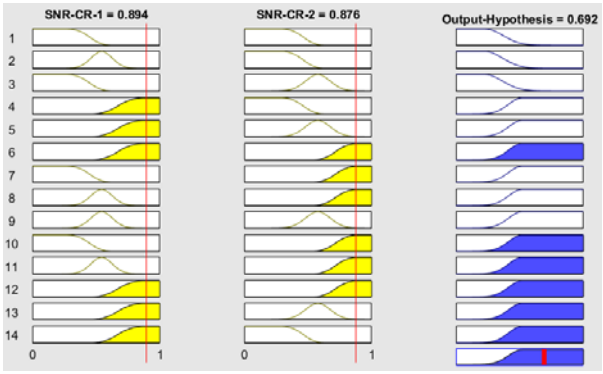


(d)



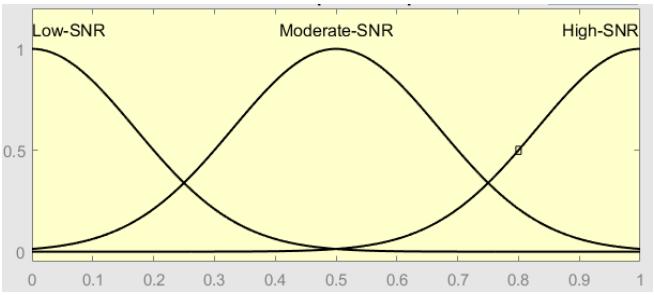
(e)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

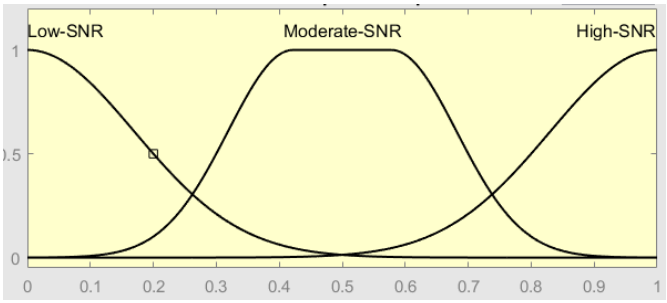


(f)

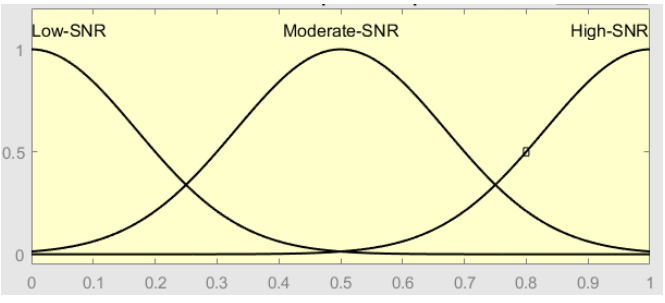
Figure 7. Results of two CRs connected to FC



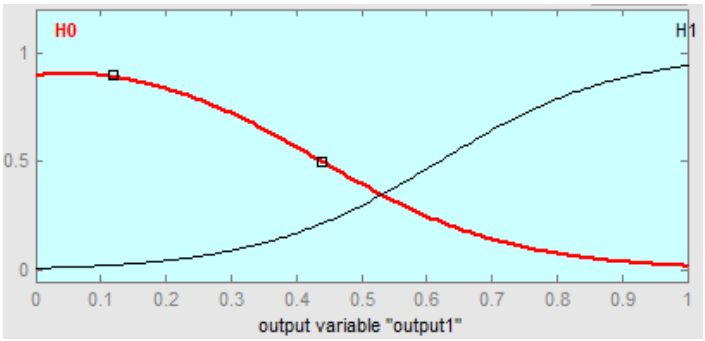
(a) SU-1



(b) SU-2



(c) SU-3



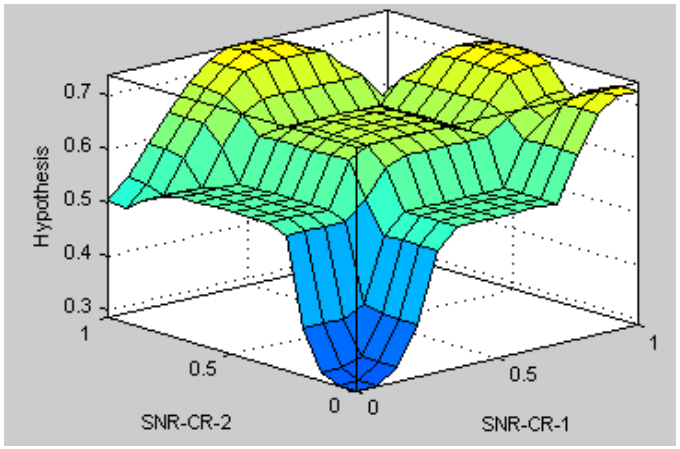
(d) Output of the Fuzzy system

Figure 8. Membership functions of inputs and output

Next we consider the fuzzy system of Fig. 4 for the case of 3 CRs, the corresponding input and output MF are shown in Fig. 8(a)-(d). There are 3^3 = 27 rules of 'and' and that of 27 for 'or'. Among then only 21 rules are shown in Fig. 9(a) and among several surface plots only 3 are shown in Fig. 9(b)-(d).

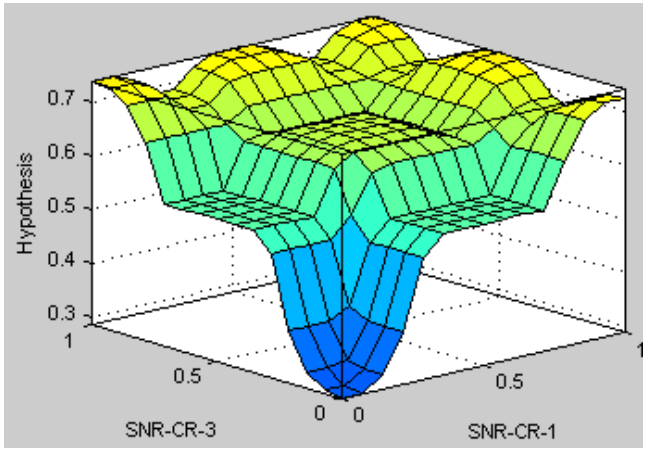
1. If (SNR-CR-1 is Low_SNR) and (SNR-CR-2 is Low_SNR) and (SNR-CR-3 is Low_SNR) then (output1 is H0) (1)
2. If (SNR-CR-1 is Moderate_SNR) and (SNR-CR-2 is Low_SNR) and (SNR-CR-3 is Low_SNR) then (output1 is H0) (1)
3. If (SNR-CR-1 is Low_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Low_SNR) then (output1 is H0) (1)
4. If (SNR-CR-1 is Low_SNR) and (SNR-CR-2 is Low_SNR) and (SNR-CR-3 is Moderate_SNR) then (output1 is H0) (1)
5. If (SNR-CR-1 is Moderate_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Low_SNR) then (output1 is H0) (0.4)
6. If (SNR-CR-1 is Moderate_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Low_SNR) then (output1 is H1) (0.6)
7. If (SNR-CR-1 is Moderate_SNR) and (SNR-CR-2 is Low_SNR) and (SNR-CR-3 is Moderate_SNR) then (output1 is H1) (0.6)
8. If (SNR-CR-1 is Moderate_SNR) and (SNR-CR-2 is Low_SNR) and (SNR-CR-3 is Moderate_SNR) then (output1 is H0) (0.4)
9. If (SNR-CR-1 is Low_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Moderate_SNR) then (output1 is H0) (0.4)
10. If (SNR-CR-1 is Low_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Moderate_SNR) then (output1 is H1) (0.6)
11. If (SNR-CR-1 is Moderate_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Moderate_SNR) then (output1 is H1) (0.8)
12. If (SNR-CR-1 is Moderate_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Moderate_SNR) then (output1 is H0) (0.2)
13. If (SNR-CR-1 is Moderate_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is High_SNR) then (output1 is H1) (1)
14. If (SNR-CR-1 is Low_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is High_SNR) then (output1 is H1) (1)
15. If (SNR-CR-1 is Low_SNR) and (SNR-CR-2 is Low_SNR) and (SNR-CR-3 is High_SNR) then (output1 is H1) (0.8)
16. If (SNR-CR-1 is Low_SNR) and (SNR-CR-2 is High_SNR) and (SNR-CR-3 is High_SNR) then (output1 is H1) (1)
17. If (SNR-CR-1 is High_SNR) and (SNR-CR-2 is High_SNR) and (SNR-CR-3 is High_SNR) then (output1 is H1) (1)
18. If (SNR-CR-1 is High_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is High_SNR) then (output1 is H1) (1)
19. If (SNR-CR-1 is High_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Moderate_SNR) then (output1 is H1) (1)
20. If (SNR-CR-1 is High_SNR) and (SNR-CR-2 is Moderate_SNR) and (SNR-CR-3 is Low_SNR) then (output1 is H1) (0.8)
21. If (SNR-CR-1 is High_SNR) or (SNR-CR-2 is High_SNR) or (SNR-CR-3 is High_SNR) then (output1 is H1) (1)

(a) Few rules of SNR model

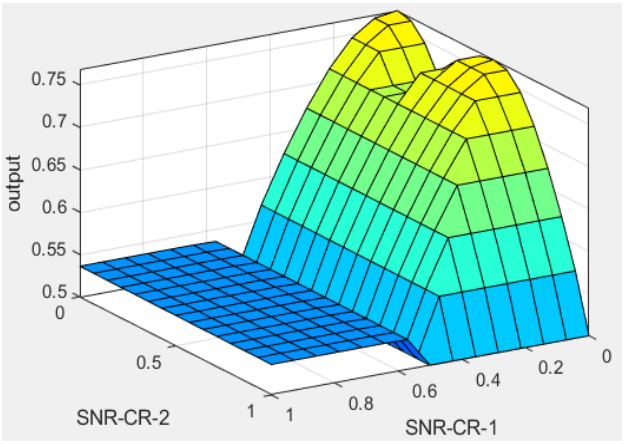


(b)

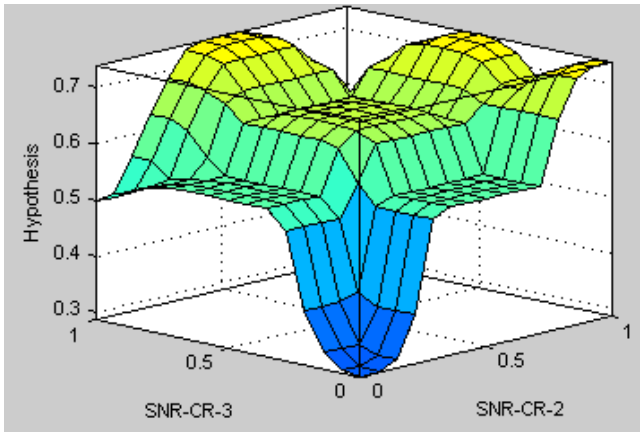
https://sites.google.com/site/ijcsis/
ISSN 1947-5500



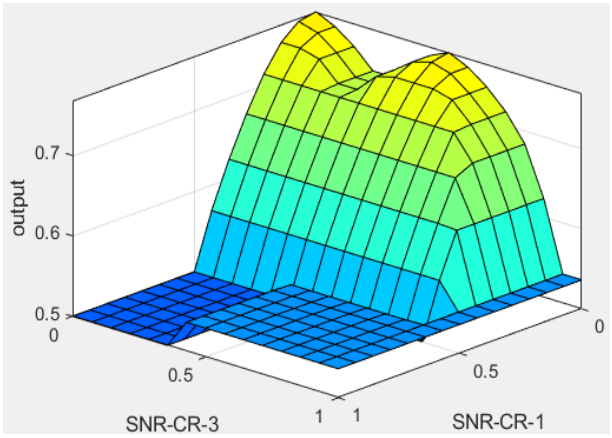
(c)



(b)



(d)



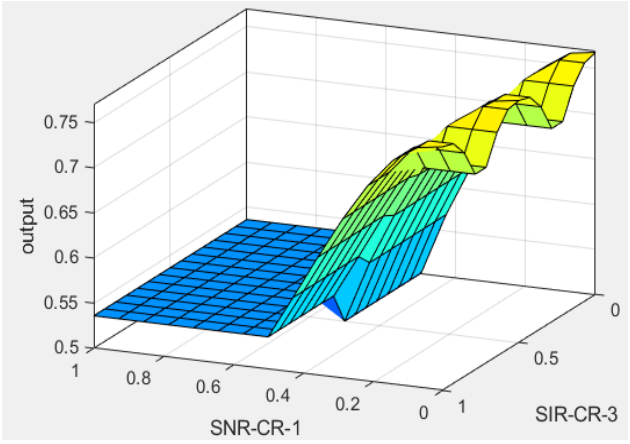
(c)

Figure 9. The surface plot of SNR two CRs against hypothesis H_1

Next we consider the second model of SNR-SIR combination of Fig. 5. Few rules of six variables: SNR-CR-1, SNR-CR-2, SNR-CR-3, SIR-CR-1, SIR-CR-2 and SIR-CR-3 are shown below and few surface plots are shown in Fig. 10 (a)-(g).

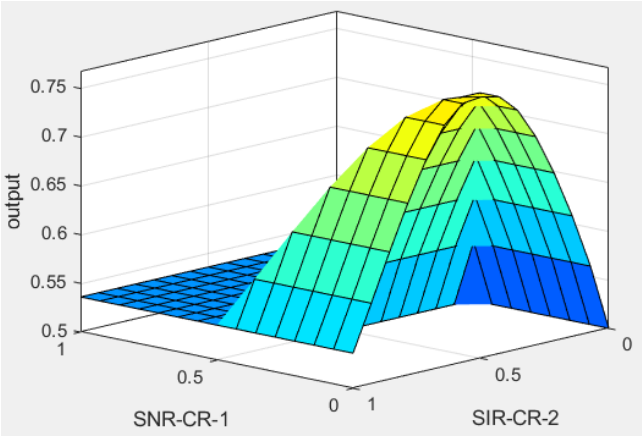
1. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H0) (1)
2. If (SNRCR-1 is MSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H0) (1)
3. If (SNRCR-1 is HSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H1) (1)
4. If (SNRCR-1 is LSNR) and (SNRCR-2 is MSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H0) (1)
5. If (SNRCR-1 is LSNR) and (SNRCR-2 is HSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H1) (1)
6. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is MSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H0) (1)
7. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is HSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H1) (1)
8. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is MSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H0) (1)
9. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is HSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H1) (1)
10. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is MSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H0) (1)
11. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is HSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H1) (1)
12. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is MSIR) then (Output-Hypothesis-CRN is H0) (1)
13. If (SNRCR-1 is LSNR) and (SNRCR-2 is LSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is HSIR) then (Output-Hypothesis-CRN is H1) (1)
14. If (SNRCR-1 is MSNR) and (SNRCR-2 is MSNR) and (SNRCR-3 is LSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H1) (0.3)
15. If (SNRCR-1 is MSNR) and (SNRCR-2 is MSNR) and (SNRCR-3 is MSNR) and (SIRCR-1 is LSIR) and (SIRCR-2 is LSIR) and (SIRCR-3 is LSIR) then (Output-Hypothesis-CRN is H1) (0.8)

(a) Few rules of SNR-SIR model

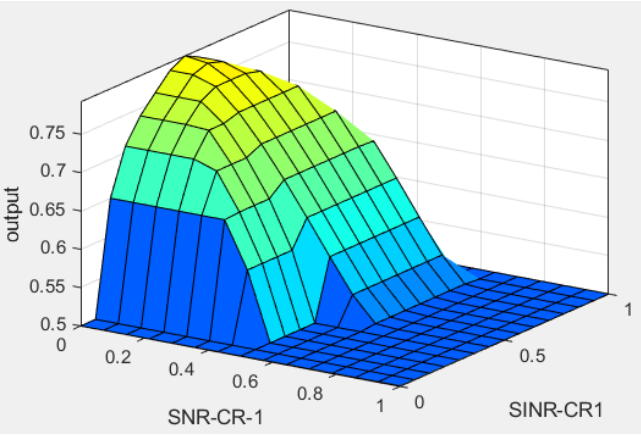


(d)

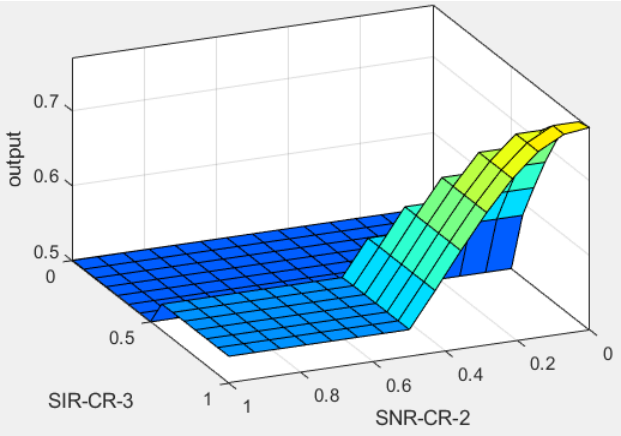
https://sites.google.com/site/ijcsis/
ISSN 1947-5500



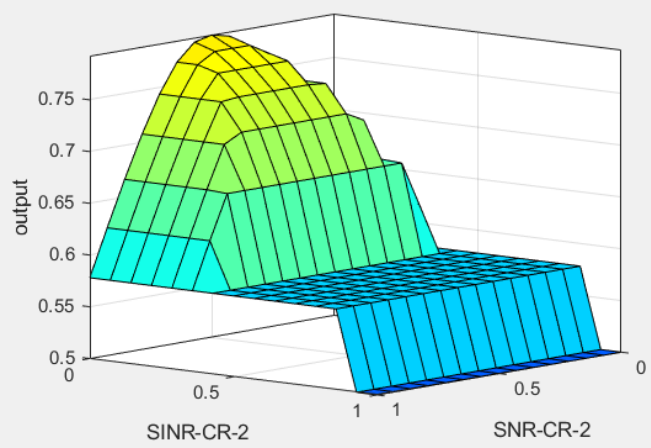
(e)



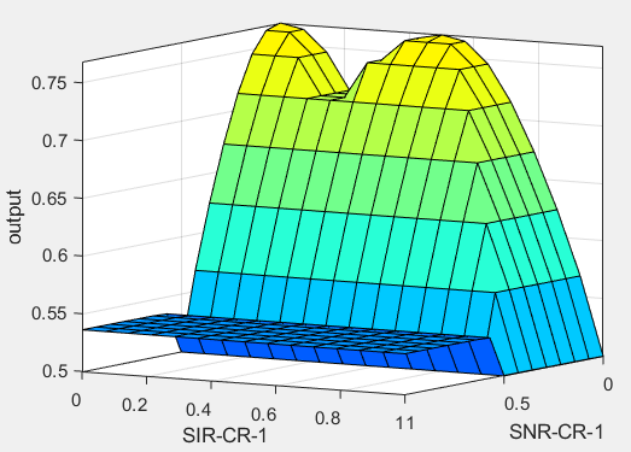
(a)



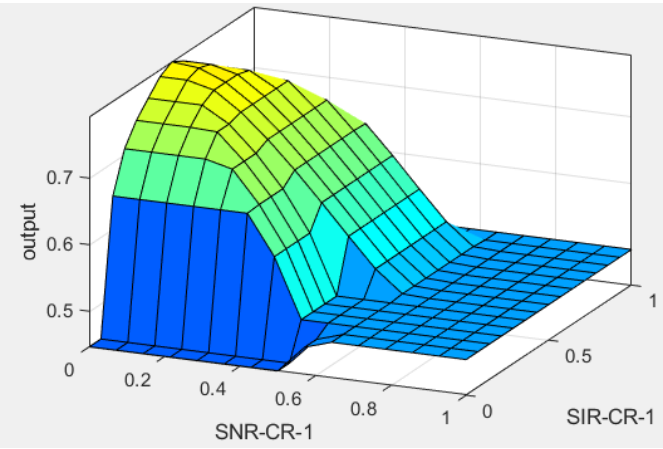
(f)



(b)



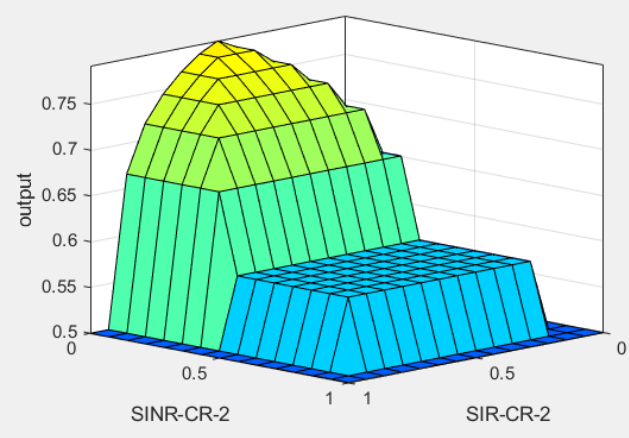
(g)



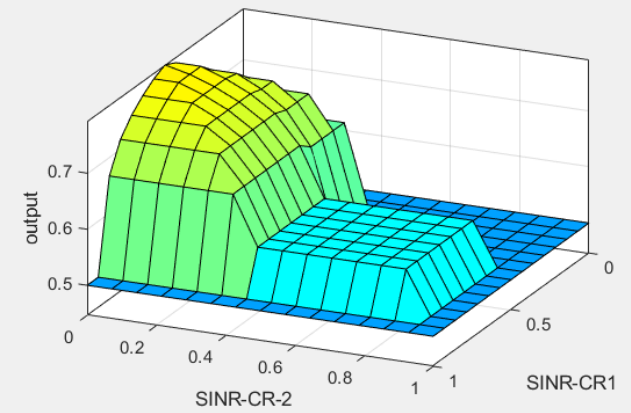
(c)

Figure 10. Surface plot of SNR and SIR combination

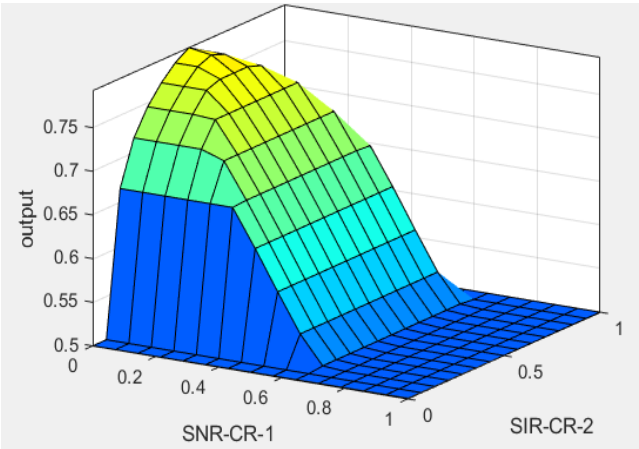
https://sites.google.com/site/ijcsis/
ISSN 1947-5500



(d)



(e)



(f)

Figure 11. Surface plot of SNR, SINR and SIR combination

Finally the combined SNR, SIR and SINR model of Fig. 6 is used and few surface plot of the system are shown in Fig. 11. The de-fuzzyfication is done by centroid method and results show that the slope of surface plot of first model is found steeper compared of second or third model. The decision is sharp in first model with possibility of error in spectrum sensing but the process time is low whereas in the second or third method the variation of possibility of detection is smooth hence less possibility of error at the expense of process time.

V. CONCLUSIONS

In this paper we only deal with three different cases of combined SNR, SINR and SIR model to detect output hypothesis. The surface plot reveals the performance of the FC. In real life a lot of parameters governs spectrum detection of a PU at FC: for example length of wireless link, path loss exponent, detection techniques and fading condition of the wireless channel etc. In this paper we model the fuzzy system under awgn only but the work can be extended under small scale fading environment where we have to include another MF of fading even more MF can be include under dissimilar fading of multi-hop link of CRs and FC.

REFERENCES

[1] R. Kaniezhil and C. Chandrasekar, 'An Efficient Spectrum Utilization via Cognitive Radio using Fuzzy Logic System for Heterogeneous Wireless Networks,' IEEE 2012 - International Conference on Emerging Trends in Science, Engineering and Technology, Tiruchirappalli, India, pp.300-307

[2] Mardeni. R, K. Anuar, Hafidzoh. M, M. Y. Alias, H. Mohamad and N. Ramli, 'Efficient Handover Algorithm Using Fuzzy Logic Underlay Power Sharing for Cognitive Radio Wireless Network,' IEEE Symposium on Wireless Technology and Applications (ISWTA), pp.53-56, September 22-25, 2013, Kuching, Malaysia

[3] Ejaz Ahmed, Liu Jie Yao, Muhammad Shiraz, Salman Ali and Abdullah Gani, 'Fuzzy-Based Spectrum Handoff and Channel Selection for Cognitive Radio Networks,' pp.23-28, 2013 International Conference on Computer, Control, Informatics and Its Applications, Jakarta, Indonesia

[4] Marja Matinmikko, Javier Del Ser, TapioRauma and Miia Mustonen, 'Fuzzy-Logic Based Framework for Spectrum Availability Assessment in Cognitive Radio Systems,' IEEE Journal on Selected Areas in Communications, vol. 31, no. 11, pp.2173-2184, November 2013

[5] Ying-Hong Wang and Shou-Li Liao, 'Applying a Fuzzy-based Dynamic Channel Allocation Mechanism to Cognitive Radio Networks,' 31st International Conference on Advanced Information Networking and Applications Workshops, pp.564-569, 2017

[6] Abdulraqeb Alhammadi, Mardeni Roslee and Mohamad Yusoff Alias, 'Fuzzy Logic Based Negotiation Approach for Spectrum Handoff in Cognitive Radio Network,' 2016 IEEE 3rd International Symposium on Telecommunication Technologies (ISTT), Kuala Lumpur, 28-30 Nov 2016, pp.120-124.

[7] Mariam Nabil and Mustafa El-Nainay, 'Fuzzy-based Assignment Algorithm for Channel Sensing Task in Cognitive Radio Networks,' 2016 IEEE Symposium on Computers and Communication (ISCC), pp.843 – 848, June 2016

[8] Nabil Giweli, Seyed Shahrestani and Hon Cheung, 'Selecting the Sensing Method in Cognitive Radio and Future Networks: A QoS-aware Fuzzy Scheme,' 2015 IEEE International Conference on Data Science and Data Intensive Systems, pp.497-504, 11-13 Dec. 2015.

[9] Joyrajchakraborty, J. V. Varma and Maria Erman, 'ANFIS based Opportunistic power control for cognitive radio in spectrum sharing,' 2013 International Conference on Electrical Information and Communication Technology (EICT), pp.1-6, 13-15 Feb' 2014.

[10] Risala Tasin Khan, Md. Imdadul Islam, Shakila Zaman and M.R. Amin, 'Optimum Access Analysis of Collaborative Spectrum Sensing in Cognitive Radio Network using MRC,' (IJACSA) International Journal of Advanced Computer Science and Applications, pp. 367-373, vol. 7, no. 7, 2016

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[11] Risala T. Khan, Md. Imdadul Islam and M. R. Amin, 'Traffic Analysis of a Cognitive Radio Network Based on the Concept of Medium Access Probability,' JIPS, vol. 10, no.4, pp. 602 ~ 617, DEC'2014

[12] Y. C. Liang, Y. Zeng, E. Peh and A. T. Hoang, 'Sensing-throughput tradeoff for cognitive radio networks,' *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, April 2008.

[13] Subhankar Chatterjee, Santi P. Maity and Tamaghna Acharya, 'Energy Efficiency in Cooperative Cognitive Radio Network in the Presence of Malicious Users,' IEEE Systems Journal, pp.1-10, Year: 2016

[14] S. Rajasekaran and G.A. Vijayalakshmi Pai, 'Neural Networks, Fuzzy Logic, and Genetic Algorithms: Synthesis and Applications,' Eastern Economy Edition; Prentice Hall of India, 2010

[15] Tallataf Rasheed , Adnan Rashdi and Ahmad Naeem Akhtar, 'Cooperative spectrum sensing using fuzzy logic for cognitive radio network,' 2018 Advances in Science and Engineering Technology International Conferences (ASET), pp.1-6, April 2018, Abu Dhabi, United Arab Emirates

[16] Mardeni Roslee , Abdulraqeb Alhammadi, Mohamad Yusoff Alias, Khairil Anuar and P. U. Nmenme, 'Efficient handoff spectrum scheme using fuzzy decision making in cognitive radio system,' 2017 3rd International Conference on Frontiers of Signal Processing (ICFSP), pp.72-75, 6-8 Sept. 2017, Paris, France

AUTHORS PROFILE

Md. Abul Kalam Azad has completed his Bachelor of Science in Computer Science & Engineering from Jahangirnagar University, Bangladesh and Master of Science in Information Technology from Royal Institute of Technology (KTH), Sweden. Currently, Mr. Azad is working as a Professor in the department of Computer Science & Engineering, Jahangirnagar University, Bangladesh. His research interest includes wireless networks, particularly in wireless sensor networks, Ad-Hoc networks, and mobile cognitive networks.

Md. Imdadul Islam has completed his B.Sc. and M.Sc Engineering in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 1993 and 1998 respectively and has completed his Ph.D degree from the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh in the field of network traffic in 2010. He is now working as a Professor at the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. Previously, he worked as an Assistant Engineer in Sheba Telecom (Pvt.) LTD (A joint venture company between Bangladesh and Malaysia, for Mobile cellular and WLL), from Sept.1994 to July 1996. Dr Islam has a very good field experience in installation and design of mobile cellular network, Radio Base Stations and Switching Centers for both mobile and WLL. His research field is network traffic, wireless communications, wavelet transform, OFDMA, WCDMA, adaptive filter theory, ANFIS and array antenna systems. He has more than hundred and seventy research papers in national and international journals and conference proceedings.

Sanjit Kumar Saha has obtained his both bachelor of science and master of science degree in computer science and engineering from Jahangirnagar University, Bangladesh in 2007 and 2009 respectively. He is currently working as an assistant professor of department of computer science and engineering at Jahangirnagar University, Bangladesh. He has 8+ years of teaching experience to both undergraduate and graduate students. Saha has published five journal paper and one conference paper in the International and National journal and conferences. His current interest includes deep learning, artificial neural network, fuzzy logic and systems.

Jugal Krishna Das has completed his Ph. D. from Glushkov Institute of Cybernetics, Kiev, Ukraine, in 1993 and M. Sc. from Donetsk Polytechnic Institute, Ukraine, in 1989. Now he is working as a Professor in the department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. He is the author of 13 Journal and more than 16 International Conference papers in home and abroad. His research interests include Network Protocols, Universal Networking Language, Distributed Systems, and so on.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Data Security on Internet of Things Device Using Hybrid Encryption Models

Marsel Sampe Asang[1], Danny Manongga[2], Irwan Sembiring[3]
[1,2,3] *Department of Information System*
Satya Wacana Christian University
Salatiga, Indonesia
[1]marselsampe@gmail.com, [2]danny@staff.uksw.edu, [3]irwan@staff.uksw.edu

***Abstract*- The development of IoT in various sectors causes new security issues. There have been many cases of breaking data and data theft on the IoT system due to poor system security, especially on the IoT device. This study focuses on how to secure data on IoT device before the data is sent to the server using a hybrid encryption model. This hybrid encryption model is a combination of AES and ECDH algorithms. Firstly this model is attempted to be implemented on RaspberryPi device. An analysis of the computational load and communication load of the encryption process is perform to measure effective and efficient of this model. The last is perform a simulation of MITM attacks to find out how well the effectiveness of this model. The results of this study shows that this encryption model is suitable enough for use on IoT device without having to overload the memory usage of device.**

***Keywords* :** *IoT Security, Encryption, Hybrid Encryption*

I. INTRODUCTION

Internet of Things is a future communication technology; every object around us will be embedded with microcontroller to make the object able to communicate with other and also with user [1]. IoT uses the internet for its connectivity so it can provide real-time information without distance limitation and human intervention, often called machine-to-machine communication.

With the growing use of IoT causes data security issues. Common types of attacks on IoT are Botnet, Man-In-The-Middle, Data & Identity Theft, Social Engineering, and Denial of Service [2]. Ensuring data security in the system is one of the biggest challenges in building an IoT system. In general, the IoT framework consists of 3 layers: sensor layer, network layer and application layer [3]. Each layer has its own security protocol in order to maintain the integrity, availability and confidentiality of the data.

The security issue on IoT is an old thread. In 2014, a security consultant conducted research on phishing and spam crimes in emails; they concluded that 25 percent of the total 750,000 spam emails were sent from smart device that connected to the internet. They take advantage of a security hole in the device and use it for Botnet attack [4]. Other related studies were also conducted in the same year, some security experts demonstrated how to attack the network used to turn on automatic lights and get the home owner's Wi-Fi username and password. [5] There are still many reports and other research related issues IoT security, but 2 (two) studies above have shown that IoT device are still vulnerable to various types of attacks, especially data theft.

The IoT device has a simple way of working; the embedded microcontroller on the device collects data from the connected sensors. Microcontrollers are also responsible for sending or receiving data from servers over the internet. The data is plaintext; the format of the data can still be read by anyone and can be a security hole for the perpetrators of crime. For example, one can do a man-in-the-middle attack to intercept the network between the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

device and server in order to read or manipulate the transmitted data. Such cases have occurred in the samsung smart fridge system which can be hacked using man-in-the-middle attacks and the offender can steal the gmail credential of the system [6].

One of the solutions for securing IoT data is using data encryption method. A compact lightweight encryption model is required in order not to overload the device's memory but keep it secure. Generating a computed lightweight encryption model is not an easy task, in fact that strong encryption definitely requires a long computation process.

This research focuses on designing, implementing and analyzing hybrid encryption models that are safe, computationally light and suitable for use in IoT device. Security discussed is more specific on the confidentiality of data, especially on the process of sending data from device to server. After designing, this encryption model is tried to be implemented on IoT device and then analyzed its performance. The analysis starts from measuring the computational load and communication load of the encryption process, and the results are compared with similar algorithms to measure its performance. The final stage is tested by Man-in-the-middle attack on the device to determine the effectiveness of this encryption model.

II. Literature Review

A. Related Research

There are many studies conducted relevant to the data security protocol on IoT device. This section presents the results of the literature Review on studies related to the title of this study.

Ngyue et al [7] conducted a comparative analysis of some types of communication security protocols on the IoT system; this research is specific to key distribution mechanisms. Similarly, O'Neill's study [8] conducted a comparative analysis of some types of security protocols on IoT device. Several other studies focus on developing and modifying existing methods. Al-Haija et al [9] implemented the TinyRSA method on arduino device to perform data encryption. Zhang et al [10] tried to integrate the microcontroller with peripheral device to perform large cryptographic computations. Yao et al [11] implements the No-Pairing ABE method based on the ECC algorithm scheme to secure data on IoT device. Stergiou et al [12] presented a data security model on the integration of IoT and Cloud Computing systems. Guicheng and Zhen [13] implements the ECC algorithm on IoT device for data authentication processes. Badra and Zeadally [14] presented a compact lightweight data security model on smart grid systems using the ECDH algorithm.

Table I is a summary of the studies described in the previous paragraph. The table briefly described the problems studied, the methods used, the results of research and the weakness of the method used.

TABLE I
SUMMARY OF THE LITERATURE REVIEW

Year and Author	Problem	Method	Result	Method Weakness
2015,	Comparative analysis of	Asymetric key schemes	Asymmetric encryption	Big computation

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Ngyue et al [7]	some types of communication security protocols of IoT, specific to key distribution mechanisms	Symmetric key pre-distribution schemes : a) Offline key distribution; b) Server-assisted key distribution	methods are more suitable for use in IoT device if optimized appropriately, and also the presence of a third party significantly impact on the distribution of the encryption key.	a) Offline key distribution: The encryption key is easy to know because it is embedded in the device; b) Server-assisted key distribution: Third-party service providers must be trusted.
2016, O'Neill [8]	Comparative analysis of various types of IoT device security protocols	Public Key Cryptography	In terms of the future security, Quantum-safe cryptography and PUF methods are the most appropriate methods to implement.	Large computation
		Quantum-Safe Cryptography (post-quantum)		Not practical and more complex and ineffective for small device.
		Physical Uclonable (PUF)		More suitable for identification and authentication.
2014, Al-Haija et al [9]	Implementation of TinyRSA 32-bit method on arduino device for data encryption	RSA 32-bit	The result of TinyRSA 32-bit implementation on arduino device runs well for encryption or decryption process.	The key length of only 32-bit is still classified as unsafe.
2013, Zhang et al[10]	Integrate microcontroller with peripheral device (Java card) for big cryptographic computation	RSA 1024-bit	The device is able to perform 1024-bit RSA encryption within 82.2 milliseconds.	The microcontroller is completely dependent on the peripheral device (Java card) for encryption
2014, Yao et al [11]	Implementation of ABE No-Pairing method based on ECC method scheme for IoT data security	Attribute-Based Encryption (ABE) with no bilinear pairing	This method is suitable to be implemented on small device because it is efficient in computing and light communication load.	Not flexible in resetting attribute values and only suitable for single-authority applications.
		ECC		Low scalability
2018, Stergio et al [12]	Data security model on IoT system integration and Cloud Computing	AES 128-bit	The model presented ensures the confidentiality of data between device and cloud computing so that device can send data via the HTTP protocol.	-
		RSA 2048-bit		Big computation
2013, Guicheng and Zhen [13]	IoT device authentication (RFID and WSN) using ECC	ECC	ECC is suitable to be implemented on IoT device due to efficiency and shorter key lengths. ECC 162-bit security is equivalent to 1024-bit RSA	The resulting ciphertext size is greater than the RSA method, so it affects bandwidth usage.
2017, Badra and Zeadally [14]	Model of lightweight computing data security on smart grid system	ECDH	The model presented is able to overcome various types of attacks such as: forgery and injection data, man-in-the-middle, known-session-key.	Big computation

In accordance with the literature review process conducted, can be noted the results and weaknesses of each method studied. It can be concluded that the most optimal method for securing data on public networks is a combination of symmetric encryption methods and asymmetric encryption. Both methods are able to overcome the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

weaknesses of each. Symmetric encryption is suitable and fast in the process of encryption and decryption of data, while asymmetric encryption fits in the secret key distribution process.

The results of this literature review will be used to determine the type of algorithm to be used in this study. The symmetric encryption method will use the AES algorithm because it is an encryption standard defined by NIST. The asymmetric encryption method uses a specific ECC algorithm on the ECDH protocol for the secret key exchange process. The reason for using ECC rather than RSA is because RSA requires a larger computation process and also the security of its encryption depends on the key length used.

B. Advanced Encryption Standard

The Advanced Encryption Standard (AES) is the standard encryption algorithm specified by NIST in 2001. AES uses the Rijndael algorithm to replace the DES (Data Encryption Standard) algorithm whose validity period has ended due to security factors [15]. AES is a type of block cipher encryption with symmetric keys. Each cipher has a 128-bit size, with key sizes 128, 192, and 256-bit each. There are 10, 12, or 14 rounds in AES; the number of this cycle corresponds to the key size used. For example, each 128-bit plaintext input is fed into a 4x4 byte square state. This state is XORed with a key and is subsequently processed 10 times by substitution process, transformation linear and round key addition. Ciphertext will be obtained in the final process.

C. Elliptic Curve Cryptography

Elliptic Curve Cryptography (ECC) is an asymmetric encryption algorithm (public key) was developed in 1985. The working principle of ECC uses elliptic curve calculations, so it differs greatly from other asymmetric encryptions that utilize very large numbers. These algorithms are often used on embedded and wireless systems, and are also popularly used for digital financial services industry signatures [15]. ECC advantages over other asymmetric encryption such as RSA and ElGamal, which is shorter key length. The ECC security level with a 160-bit key is equivalent to RSA with 1024-bit keys [11] [13] [15]. The ECC key length is relatively short making this algorithm suitable for use on small device that have limited resources. ECC algorithm has been widely implemented in various protocols, for example the Elliptic Curve Diffie-Hellman (ECDH) protocol, this protocol is designed specifically to handle the exchange of secret keys on the public network.

III. Research Contribution

This research describes a hybrid encryption model that combines the AES and ECDH algorithms to generate secure, lightly computed and suitable encryption on IoT device. This model is attempted to be implemented on the RaspberryPi device to test and analyze its computing and communications loads.

IV. Proposed Model

Commonly the data sent from the device to the server is in plaintext form, can be read by anyone. To secure the data, it needs to be encrypted so that unauthorized parties will not be able to read it. The encryption process will convert the plaintext data into ciphertext. Symmetric encryption is the most suitable method used in small devices, this method is faster and efficient computation load than asymmetric encryption. The disadvantage of symmetric encryption is that the encryption key must be embedded in the physical memory of the device, thus causing the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

vulnerable encryption key to be known by unauthorized parties. Embedding the encryption key on the physical memory of the device is a poor technique to implement.

This research describes hybrid encryption models where symmetric encryption keys will be generated based on agreement between server and device when interconnected. In this model, the device can only encrypt the data using the secret key.

A. Hybrid Encryption Model

The hybrid encryption model ensures the exchange of secret keys between server and device, so the device is able to perform symmetric encryption without having to plant secret keys in its physical memory. The working mechanism of this model begins when the device is connected to the server. Server and connected device will exchange public keys, and then together generate a secret key. This secret key will be used to perform data encryption.

The hybrid encryption described in this study uses the ECDH algorithm for key distribution process and AES algorithm for data encryption process. ECDH generates secret keys known only to server and device. While AES secures the data on the device before it is sent to the server. Both algorithms use a safe key length, AES with 256-bit key length and ECDH with 256-bit key length.

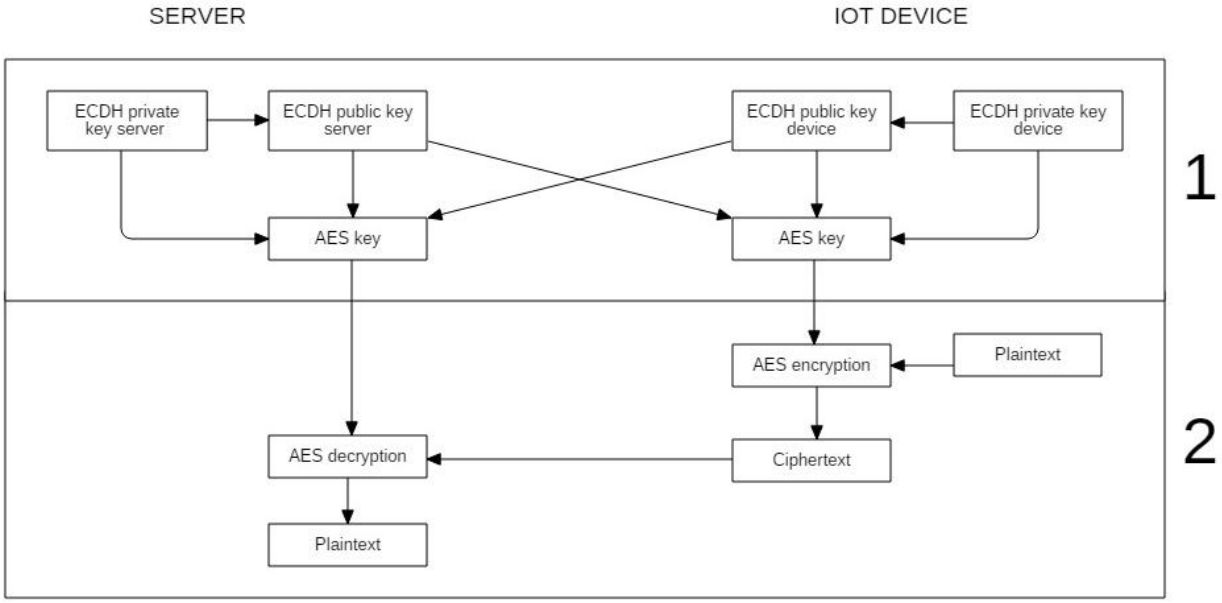


Figure 1. Flowchart of hybrid encryption model

Figure 1 is a flowchart of the workings of the hybrid encryption model in this study. This model consists of 2 main processes; the key distribution process and the process of securing the data.

The key distribution process uses the ECDH algorithm. In this process, server and device create a public key based on their private key. Then the server and device will exchange public keys, and with the public key will generate a secret key known only to server and device. This process is only done once just after the device is connected to the server.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The process of securing the data uses AES algorithm. The plaintext data in the device will be encrypted resulting in ciphertext. Ciphertext sent to the server and the server will decrypt it back to plaintext. This process is always done if the device will sends the latest data to the server.

V. Performance Encryption Analysis

This research method uses experimental method that is by doing measurement of computation load and communication load required by encryption process. Measurements are performed on two main processes, namely the key distribution process and the process of securing the data. This test focuses on the device only and is not performed on the server because the encryption load on the server depends on the hardware specification, so the value is inconsistent to be measured.

The computational load is the length of time it takes to perform a process, its value is calculated in milliseconds. The communication load is the amount of data generated from a process that will be sent from server to device or vice versa over the internet network, its value is calculated in byte units.

After measuring the computational load and the communication load, the ECDH algorithm will be compared to similar algorithms. The purpose of doing this comparison is to measure how well the ECDH algorithm performs when performing key distributions on IoT device.

A. Specification of IoT Device

The IoT device discussed in this study uses RaspberryPi 3 model B with hardware specifications Quad Core 1.2GHz CPU, 1GB RAM, 8GB SD Card and software specification using Windows IoT Core OS.

B. Measurement of Computation and Communication Load

Measurements are performed on two main processes that occur in the device, namely the distribution of the encryption key and the process of securing the data before it is sent to the server. The measurement techniques performed on the two main processes are quite simple. The computational load value is calculated using the Stopwatch feature from dot Net framework; while the value of the communication load is calculated based on the size of the data packets recorded using the WireShark software (packet capturing software).

Stopwatch feature is perfect for measuring the computation time of a process. The stopwatch is embedded in the program code, and automatically runs and stops at the specified point. Figure 2 shows the pseudo code of the secret key distribution process using the ECDH algorithm embedded in the Stopwatch feature to measure the length of process computation.

```
1.  Start device
2.  Initialize
3.  Start Stopwatch
4.  s1 = Generate private key()     // generate device private key
5.  p1 = Generate_public_key()      // generate device public key
6.  Send_to_server(p1)              // send device public key to server
7.  p2 = Receive_from_server()      // receive server public key
8.  key= Generate_secret_key(p1,p2) // generate the secret key
9.  Stop Stopwatch
10. Do_encryption(data,key)
```

Figure 2. Pseudo code of secret key distribution process using ECDH algorithm

Table II describes the measurement results of the computational load of the device. The computational load value obtained is dynamic and changing according to the performance of the device processor. Mean value is the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

average of computational load, while the value of Stdev is the standard deviation of computational load, detail calculation of Mean and Stdev value is presented in table III.

TABLE II
THE MEASURING RESULT OF THE DEVICE COMPUTATIONAL LOAD

Process	Algorithm	Data processed	Computational load		Notes
			Mean	Stdev	
Key distribution	*ECDH* (256-bit)	32 *bytes*	140.6 ms	48.601 ms	Generating 32 bytes encryption keys
Secure the data	*AES* (256-bit)	32 *bytes*	307.9 ms	72.787 ms	Encrypt 32 bytes messages.

TABLE III
CALCULATION OF MEAN AND STDEV ON COMPUTATIONAL LOAD VALUES

Experiment	Key Distribution (*ECDH*)	Secure the data (*AES*)
1	179 ms	303 ms
2	101 ms	345 ms
3	94 ms	343 ms
4	183 ms	204 ms
5	92 ms	380 ms
6	93 ms	370 ms
7	199 ms	206 ms
8	94 ms	357 ms
9	184 ms	361 ms
10	187 ms	210 ms
Mean	**140.6 ms**	**307.9 ms**
Stdev	**48.601 ms**	**72.787 ms**

Table IV, describes the measurement results of the communication load on the device. The value of the obtained communication load is static according to the size of the processed data. This is different from the computational load whose value is always changing.

TABLE IV
THE MEASURING RESULT OF THE DEVICE COMMUNICATION LOAD

Process	Algorithm	Data processed	Comunication load	Notes
Key distribution	*ECDH* (256 bit)	32 *bytes*	376 bytes	Generating 32 bytes encryption keys
Secure the data	*AES* (256 bit)	32 *bytes*	108 bytes	Encrypt 32 bytes messages.

C. Comparison between ECDH and RSA in key distribution processes

A comparison with similar algorithms aims to measure how well the performance of the ECDH algorithm performs the key distribution of IoT device. The ECDH algorithm is compared with RSA, because the RSA algorithm is one of the popular algorithms used to perform the distribution of secret keys.

Figure 3 shows a comparative load comparison chart between the ECDH and RSA algorithms when performing the secret key distribution process on server and device. From the graph, it is seen that the performance of ECDH is much better than RSA when doing the key distribution.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

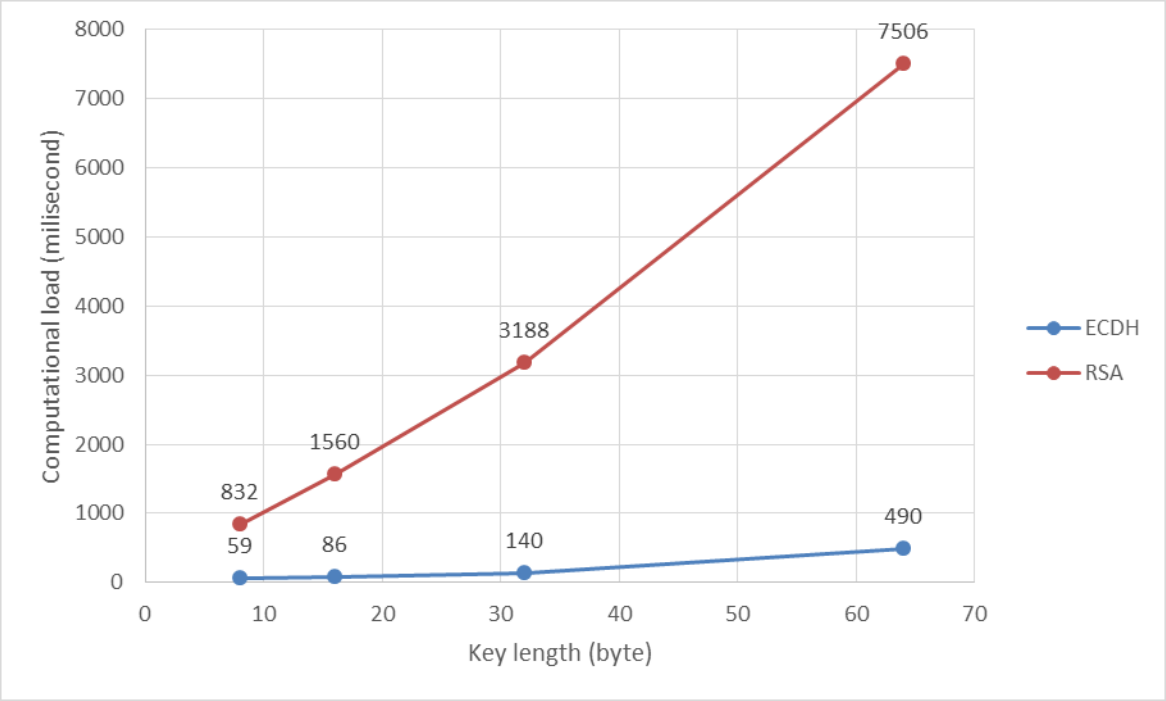


Figure 3. Comparison of computational loads between ECDH and RSA algorithms on performing key distributions

Based on the computational load and communication load measurement and the result of comparison between similar algorithms, it can be concluded that the performance of this model is good enough to be used on IoT device.

VI. Attack Simulation

The attack simulation aims to see the effectiveness of the hybrid encryption model that is implemented on the device. The attack will try to read the original data sent from the device to the server via http request. The original data is data that has not been secured using encryption, such as the data in http request in figure 4.

```
POST /savedata.php HTTP/1.0
Host: iot.marselsampeasang.web.id
Content-type: application/x-www-form-urlencoded
Content-Length: 18

this is dummy data
```

Figure 4. HTTP Request with unencrypted content

The type of attack to be tested is man-in-the-middle (MITM) attack. Attacks by intervening the network between the device and the server so that the attacker can read or manipulate the data sent. MITM attacks can be classified into 2 (two) types: passive attacks and active attacks. Passive attacks aimed at viewing and analyzing successfully tapped data, these attacks do not manipulate data. Active attacks aim to change and manipulate data sent between device and server.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

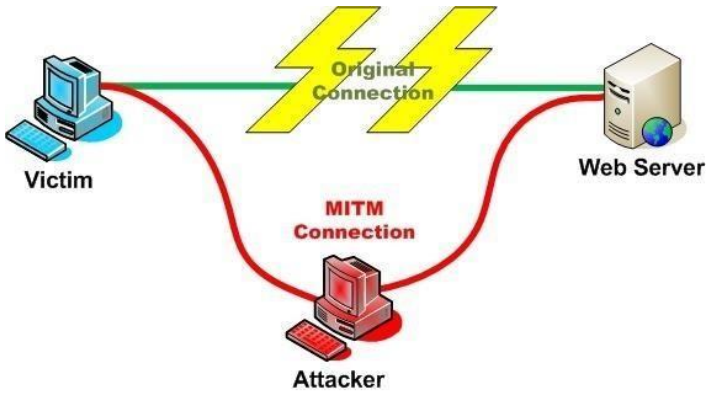


Figure 5. Man-in-the-middle (MITM) attack

Passive attack simulation aims to read and analyze data successfully intercepted from the device. The simulation uses WireShark software (packet capturing software). The software will record every http request from the device leading to the server. Figure 6 shows the http request successfully recorded, in the request the data sent is encrypted and can no longer be read. The result of this attack fails to read device data because it must be decrypted using a secret key.

```
POST /savedata.php HTTP/1.0
Host: iot.marselsampeasang.web.id
Content-type: application/x-www-form-urlencoded
Content-Length: 65

I3mMDvgCRri11UOn32XMZ+V4ZM6Z4cf9RwfTTvuhK+sZO35nOCDLGYbXvBLamisb
```

Figure 6. HTTP Request with encrypted content

Active attack simulation aims to generate the secret key that the device uses for the encryption process. This simulation uses the Address Resolution Protocol (ARP) poisoning technique on the device router, as shown in Figure 7. When the device performs the secret key distribution process with the server, the device receives a fake server public key with consideration of the key distribution between the attacker and the device.

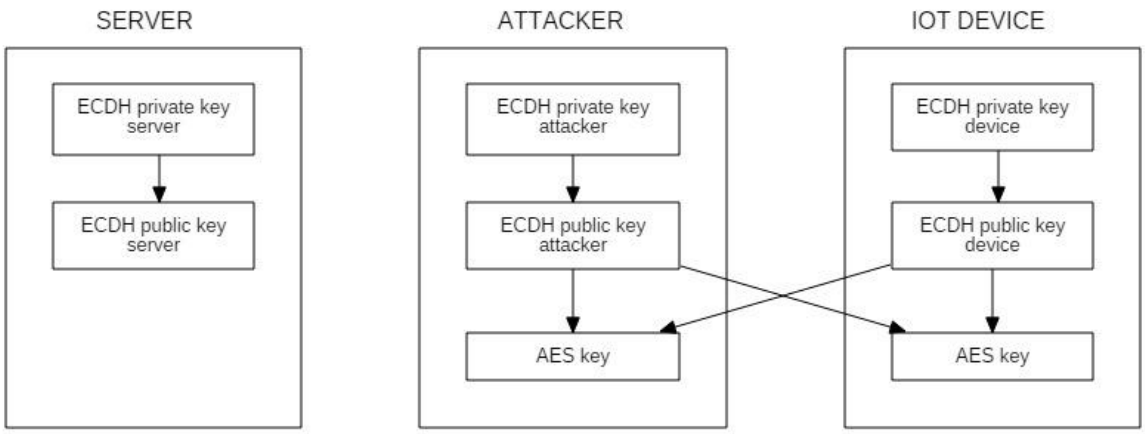


Figure 7. Simulation ARP poisoning attack to obtain the secret key

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The above technique is successful and the distribution of the secret key occurs only between the attacker and the device. The device will encrypt data using the same secret key as the attacker. The result of this attack successfully reads the encrypted data of the device.

MITM attack simulation results show that hybrid encryption model in this research is safe against sniffing attack, but still susceptible to ARP poisoning attack. Security at the network layer needs to be improved to cope with such attacks

VII. Limitations of the Proposed Model

Although the performance generated by this encryption model is quite good, there are some limitations that need to be improved in the future.

1) The secret key for encryption is not the key that is completely derived from the server. This secret key is generated by the ECDH algorithm based on the value of the server's public key and device.
2) Each device must have a different private key, so the public keys and secret keys generated by each device will vary. The main purpose of distinguishing private keys is to prevent unauthorized parties from generating the same secret key to all device.
3) Private key and public key management on the device are static.
4) This model is still vulnerable to MITM attacks when network layer security is poor.

VIII. Conclusion

This study describes hybrid computing models that are computed lightly and are suitable for securing data on IoT device. This model uses a combination of ECDH and AES algorithms. This model is attempted to be implemented on RaspberryPi device to test its encryption performance. Performance is tested based on the value of computational load and the required communication load of 2 (two) main processes ie the key distribution process and the process of securing the data. The analysis results show that this model is suitable for use on IoT device without having to burden the device memory. MITM attack simulation are also conducted to see the effectiveness of this model. Some limitations of this model are also presented in order to be addressed in the future. Advice on future development is to develop better key management on the device side, and also test server scalability when accessed by multiple device simultaneously.

References

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
[2] L. Toms, "5 Common Cyber Attacks in the IoT - Threat Alert on a Grand Scale," *GlobalSign Blog*, 2016. [Online]. Available: https://www.globalsign.com/en/blog/five-common-cyber-attacks-in-the-iot/. [Accessed: 03-Jul-2016].
[3] T. Li, "Principle, Framework and Application of Internet of Things," *Appl. Mech. Mater.*, vol. 303–306, pp. 2144–2148, 2013.
[4] Proofpoint, "Proofpoint Uncovers Internet of Things (IoT) Cyberattack," 2014. [Online]. Available: https://www.proofpoint.com/us/proofpoint-uncovers-internet-things-iot-cyberattack. [Accessed: 03-Sep-2017].
[5] A. Chapman, "Hacking into Internet Connected Light Bulbs," 2014. [Online]. Available: https://www.contextis.com/blog/hacking-into-internet-connected-light-bulbs. [Accessed: 03-Sep-2017].
[6] P. Partner, "Hacking DefCon 23's IoT Village Samsung fridge," 2015. [Online]. Available: https://www.pentestpartners.com/security-blog/hacking-defcon-23s-iot-village-samsung-fridge/. [Accessed: 13-Jun-2018].
[7] K. T. Nguyen *et al.*, "Survey on secure communication protocols for the Internet of Things," *Bus. Horiz.*, vol. 58, no. 6, pp. 17–31, 2015.
[8] M. O'Neill, "Insecurity by Design: Today's IoT Device Security Problem," *Engineering*, vol. 2, no. 1, pp. 48–49, 2016.
[9] Q. A. Al-Haija, M. Al Tarayrah, H. Al-Qadeeb, and A. Al-Lwaimi, "A tiny RSA cryptosystem based on arduino microcontroller useful for small scale networks," *Procedia Comput. Sci.*, vol. 34, pp. 639–646, 2014.
[10] E. P. Zhang *et al.*, "A Simple and Efficient Way to Combine Microcontrollers with RSA Cryptography," vol. I, pp. 23–25, 2013.
[11] X. Yao, Z. Chen, and Y. Tian, "A lightweight attribute-based encryption scheme for the Internet of Things," *Futur. Gener. Comput. Syst.*, vol. 49, pp. 104–112, 2014.
[12] C. Stegrious, K. E. Psannis, B.-G. Kim, and B. Gupta, "Secure Integration of IoT and cloud computing," *Futur. Gener. Comput. Syst.*, vol.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

78, no. 3, pp. 964–975, 2018.
[13] S. Guicheng and Y. Zhen, “Application of elliptic curve cryptography in node authentication of internet of things,” *Proc. - 2013 9th Int. Conf. Intell. Inf. Hiding Multimed. Signal Process. IIH-MSP 2013*, pp. 452–455, 2013.
[14] B. Mohammad and Z. Sherali, “Lightweight and efficient privacy-preserving data aggreration approach for the smart grid,” *Ad Hoc Networks*, vol. 64, pp. 32–40, 2017.
[15] N. Galbreath, *Cryptography for Internet and Database Applications*. 2002.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

MOBILE PHONE CLONING - A CONCEPTUAL REVIEW

*Onwuama T.U
Department of Computer Science
Federal University of Technology
Owerri, Nigeria
Udora09@gmail.com

Onukwugha C.G
Department of Computer Science
Federal University of Technology
Owerri, Nigeria
Onukwugha2000@yahoo.com

Odii J.N
Department of Computer Science
Federal University of Technology
Owerri, Nigeria
jnodii@yahoo.com

Nwokoma F.O
Department of Computer Science
Federal University of Technology
Owerri Nigeria
adibefrancisca@yahoo.com

Abstract

Mobile phone cloning is the practice of taking the programmed information stored in a valid mobile phone and criminally programming the same information into another mobile phone. This cell phone piracy has become more prevalent in recent times and of course a matter serious concern in the computing world since the rate at which it is used to commit crime is increasing at an alarming rate. Consequently, this paper reviews mobile phone cloning in both GSM and CDMA networks, the various strategies used by the perpetrators for mobile phone cloning, different ways of its detection and most importantly, its prevention.

Keywords: Cloning, Piracy, Mobile Phone, ESN, MIN, GSM, CDMA

1.0 Introduction

Mobile phone cloning is the process of transferring the identity of one phone to the other, the intent most of the times is to commit fraud. Mobile phone cloning also known as cell phone piracy has been taking place throughout the world since decades (Manjula & Rajanna 2015). Mobile phones are essential parts of human life. They are easy to use, efficient and economical. One can hardly do without phones these

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

days. It has also become an area of interest since a lot of revenue can be generated from it since a lot of businesses depend on it.

In mobile phone cloning, the subscriber information taken from one phone is copied onto the other with the intention of obtaining free calls. The other mobile phone becomes the exact replica of the original mobile phone like a clone. As a result, while calls can be made from both phones, only the original phone is billed.

Millions of mobile phones users, be it GSM (Global System for Mobile Communications) or CDMA (Code Division Multiple Access) run at threat of having their phones cloned. As a mobile phone user if you have been receiving enormously high bills for calls that you never placed, chances are that your mobile phone is possibly cloned. Unfortunately, the subscriber may not easily suspect that his/her phone has been cloned. But actions like call failing or anomalies in monthly bills can act as tickers (Akash et al, 2014). The cloner can set the options to ring his phone when the victim makes a call and the victim will have no idea that the cloner is listening from the cloner's own mobile. The cloner can read text message, phone book entries, look at pictures etc. Also, the cloner can dial phone numbers from his phone and a whole lot more. So when one gets huge bills, the chances are that the phone is being cloned. Cell phone cloning (Eureka, 2017) started with Motorola "bag" phones and reached its peak in the mid 90′s with a commonly available modification for Motorola "brick" phones such as the Classic, the Ultra Classic, and the Model 8000. Cloning involved modifying or replacing the EPROM in the phone with a new chip, which would allow one to configure an ESN (Electronic Serial Number) via software. The MIN (Mobile Identification Number) would also have to be changed. After successfully changing the ESN/MIN pair, the phone would become an effective clone of the other phone. Cloning only requires access to ESN and MIN pairs. And ESN/MIN pairs can be discovered in several ways: Sniffing the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

cellular network, Trashing cellular companies or cellular resellers and Hacking cellular companies or cellular resellers.

Hence since cloning most of the time only requires access to ESN and MIN of the targeted phones, and also given the porosity of our networks, it therefore becomes pertinent that phone users should be acquainted with the possibility of their phones being cloned, and at the same time used for fraud without their knowledge. This paper is therefore timely given the prevalence of phone related crimes in the world today.

2.0 Related Literature

2.1 How Mobile Phone Works

According to Akash, et al (2014), mobile phones send radio frequency transmissions through the sky on two distinct channels, one for voice communications and the other for control signals. When a mobile phone builds a call, it normally transmits its Electronic Security Number (ESN), Mobile Identification Number (MIN), its Station Class Mark (SCM) and the number called in a tiny burst of data. This burst is the short buzz you hear after you press the SEND button and before the tower catches the data. These four things are the components the cellular supplier uses to ensure that the phone is programmed to be billed and that it also has the identity of both the customer and the phone. MIN and ESN are collectively known as the ‘Pair’ which is used for the cell phone identification. When the cell site gets the pair signal, it determines if the requester is a valid registered user by comparing the requestor's pair to a cellular subscriber list. Once the cellular telephone's pair has been recognized, the cell site emits a control signal to permit the subscriber to place calls

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

at will. This practice, known as Anonymous Registration, is carried out each time the telephone is turned on or picked up by a new cell site.

2.2 Electronic Serial Number (ESN)

The unique identification number according to Eureka (2017), is embedded in a wireless phone by the manufacturer. Each time a call is placed, the ESN is automatically transmitted to the base station so the wireless carrier's mobile switching office can check the call's validity. The ESN cannot easily be altered in the field. The ESN differs from the mobile identification number, which is the wireless carrier's identifier for a phone in the network. MINs and ESNs can be electronically checked to help prevent fraud and features. Each ESN is a 32-bit number consisting of three fields: a manufacturer code (eight bits), a unique serial number (eighteen bits), and an extension (six bits). In practice, the serial number and the extension have actually been combined into one 24-bit serial number to identify each mobile unit. Under this assignment format, 256 manufacturers could be distinguished by ESN. But when this number proved insufficient, the 32-bit ESN assignment was altered to reflect a 14-bit manufacturer code and an 18-bit unit identification number.

2.3 Mobile Identification Number (MIN)

The Mobile Identification Number (MIN) (Eureka, 2017) is a number that uniquely identifies a mobile telephone subscriber. MINs are 34-bits in length. The first 10 bits are sometimes known as MIN2, while the last 24 bits are referred to as MIN1. Together they are simply known as the MIN.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

2.4 How ESN/MIN Are Detected

Cellular thieves can capture ESN/MINs using devices such as cellphone ESN reader or digital data interpreters (DDI). DDIs are devices specially manufactured to intercept ESN/MINs. By simply sitting near busy roads where the volume of cellular traffic is high, cellular thieves monitoring the radio wave transmissions from the cellphones of legitimate subscribers can capture ESN/MIN pair. Numbers can be recorded by hand, one-by-one, or stored in the box and later, downloaded to a computer. ESN/MIN readers can also be used from inside an offender's home, office, or hotel room, increasing the difficulty of detection. (Manjula and Rajanna, 2015).

2.5 How ESN/MIN Are Programmed on another Phone

To reprogram a phone, the ESN/MINs are transferred using a computer loaded with specialized software, or a “copycat” box, a device whose sole purpose is to clone phones. The devices are connected to the cellular handsets and the new identifying information is entered into the phone. There are also more discreet, concealable devices used to clone cellular phones. Plugs and ES-Pros which are about the size of a pager or small calculator do not require computers or copycat boxes for cloning. The entire programming process takes 10-15 minutes per phone. (Manjula and Rajanna, 2015)

2.6 GSM Phones

Global System for Mobile Communications (GSM) is a digital cellular phone technology based on time division multiple access (TDMA). GSM phones use a Subscriber Identity Module (SIM) card that contains user account information. Any

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

GSM phone becomes immediately programmed after plugging in the SIM card, thus allowing GSM phones to be easily rented or borrowed. Operators who provide GSM service are Airtel, Hutch etc.

In their work, (Mishra and Nilesh 2014) said that mobile service providers needed to secure their networks from attack and misappropriation of networking resources. In the attempt to achieve the goals set out in GSM of protecting access to mobile services and to protect any relevant item from being disclosed on the radio path; the GSM security protocols were developed. There are many technical constraints that are needed to be addressed when adding security to mobile communication. When authenticating against a mobile wireless network, the mobile equipment needs to be able to send from one base station to another without a loss of communication or interruption to an active connection. The requirement to roam without interruption was a major factor in development of mobile networks that would allow a user to be able to authenticate to and use all parts of the network seamlessly. A major difficulty faced by mobile networks is the ability for a user to roam from one network to another network operator which allows mobile network providers to bill foreign users and systems. The authentication protocol deployed to address these problem was the SIM based GSM protocol. In GSM networks, a mobile station is connected to visit network by several radio link to a particular base station.

2.7 CDMA Phones

CDMA is a method for transmitting simultaneous signals over a shared portion of the spectrum. There is no Subscriber Identity Module (SIM) card unlike in GSM. The digital cellular standard for CDMA is developed by Qualcomm. CDMA is a 2G mobile telecommunications standard. In CDMA, the same frequencies are allocated

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

to share multiple radios links. It is a type of multiplexing which is used to optimize the bandwidth of single channel. CDMA is a form of multiplexing, which allows several signals to optimize the available bandwidth. (Aaruni et. al, 2012)

2.8 Cloning a Mobile Phone

Each mobile phone has a specific broadcasting fingerprint in its transmitted information signal. This fingerprint is very unique for a particular number. This print does not get altered even if the user changes MIN or ESN number. The process of Cloning access ESN and MIN pair in following ways to make a success:

a. Sniffing of radio waves sniffing devices.
b. Usage of garbage of mobile phones or hacking of mobile phone service Provider Company.
c. Breach the security to gain unauthorized access in mobile companies.

2.8.1 Cloning GSM Phones

Cloning has been shown to be successful on code division multiple access (CDMA) but more difficult on the Global System for Mobile (GSM). GSM is one of the most widely used mobile telephone communication systems. However, cloning GSM phones is achieved by cloning the SIM card contained within, not necessarily any of the phone's internal data. Cloning of GSM mobile is a rare process. It is one of the reasons that make GSM phone more popular as cloning of such mobile is only possible through the cloning of SIM card inserted into it. The main reason for this is that these phones do not have ESN or MIN number. They only have IMEI number. SIM can be copied by removing the SIM card and placing a device between handset

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

and SIM card to extract KI or secret code. This process may take a few days. The process of cloning in GSM mobile phone is a tough process so it is being a research area for researchers.

2.8.2 Cloning for CDMA Mobile Phones

CDMA clone transfers all the user setting and data from original legitimate phone into fraudulent phone that is indistinguishable in make and firmware version. In CDMA mobile, the EPROM (Erasable Programmable Read-Only Memory) is replaced with a new chip with new configured ESN by the use of software. The second step is to change the MIN and to make a successful ESN/MIN pair. This pair sometimes pronounced as Mobile Equipment Identifier (MEID). The ESN/MIN is transmitted to cellular company to authenticate device into mobile network. After making this modification, the mobile phone PRL and number itself or MIN number can pave the way for fraudulent calls, as the target mobile phone is now the clone of the mobile phone from where the original ESN and MIN numbers are obtained. Cloning in CDMA mobile requires ESN and MIN pair. The figure below shows a typical pictorial view of how a phone is cloned.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

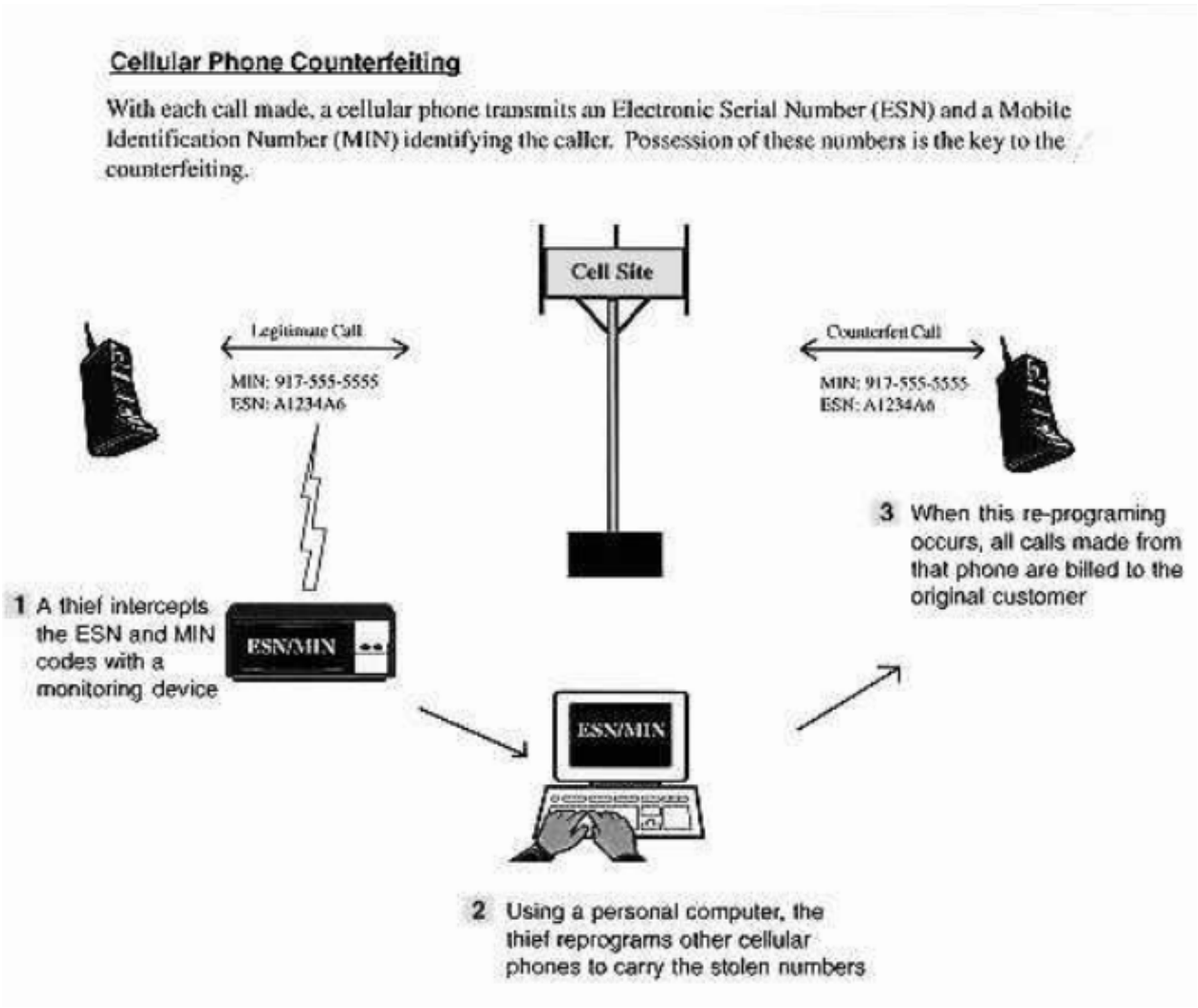


Figure 1: Cellular Counterfeiting/Cloning Fraud.

Source: (Manjush et al, 2013)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

3.0 Detection of Phone Cloning

These are the different ways of detecting a suspected cloned phone.

i. When there is duplicate Location

This is also called duplicate detection. When a service provider discovers that the same phone is used in several places at the same time. The service provider may shut down the network and wait for the legal user to respond back to the service provider. At this point, the ESN/ MIN can be reprogrammed resulting in detection of the false user. The only loophole in this system is that it is very much difficult for the service provider to trace out the duplicates.

ii. Velocity Trap

This is almost like the first one above, but in this case, if the location of the phone is continuously changing or the location is too far away from last call within the seemingly impossible amount of time, then it falls under velocity trap. For an example, if first call is made from Nigeria and another is made from Canada within sixty minutes, or if the calls are encountered from Lagos and another Owerri within three minutes, Velocity Trap is suspected.

iii. Radio Fingerprint

This is the process of identifying a cellular phone or any other device by a unique "fingerprint" that characterizes its signal transmission. The identification of a wireless device is done by the electronic fingerprint detected due to its unique radio transmission characteristics. Cellular operators use Radio fingerprinting to prevent

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

cloning of mobile phones. When a phone is cloned, it will have a similar numeric equipment identity but a different radio fingerprint.

iv. Usage Pattern

This can be called usage profiling. When the usage patterns of the users are studied, any obvious differences can be noted, and the original authenticated user is contacted. For instance, if a user is normally known for local calls and suddenly or a call is tracked immediately from foreign country, then it's possible the phone has been cloned.

v. Call Logs

Every phone keeps records or logs of calls it has been used for since purchased. Every service provider also keeps the same logs. If the logs from the service provider and the user's logs are not matched, then chances are that the phone has been cloned. Note that call logs is also known as call counting.

vi. Smart PIN Code

This is a case where the service provider assigns a smart PIN (Personal Identification Number) code to an authentic user. The authentic user will request for service privilege from service provider and temporary suspension of service before and after each call respectively. This PIN code is normally shared by authenticated user and service Provider. The encryption standards and the security algorithms, can be implemented on this PIN rather than ESN/MIN Pair.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

4.0 Suggestive Factors that indicate that a Phone has been Cloned.

i. There will be prevalent difficulty in receiving calls or voice messages.
ii. The owner of the mobile phone will notice recurrent phone calls from wrong numbers.
iii. There will also be difficulty in making calls even when there is sufficient network.
iv. Continuous receipt of line busy signals when trying to place calls
v. Receiving messages from wrong numbers.
vi. Outrageous call bills from service providers.
vii. Inconsistent bills for an unknown transaction.
viii. Anomalous transactions not initiated by the rightful phone user.
ix. Call breaching.

5.0 Cell Phone Cloning- Possible Preventive Measures

Here are possible number of ways in which one can prevent his/her phones from being cloned. They include:

i. Crosscheck your bills either daily weekly or even monthly to ensure there are no much variations in the bills. You may wish to contact your service provider.
ii. Saving of highly classified information in mobile phones should be minimized.
iii. Ensure that all your mobile phones are under insurance by the company policy.
iv. Use of password to secure your phone.
v. You really don’t have to trust anybody with your phone

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

vi. Change your phone password from time to time.

vii. Use more of GSM phones than CDMA phones if you can't be too careful as CDMA phones are more difficult to clone.

6.0 Conclusion

This paper has examined the concept of phone cloning, its consequences, and possible ways of detecting a cloned phone and also suggested a number of preventive measures for the aversion of this crime. It therefore behooves on individual phone users and business owners to take the issue of mobile phone cloning seriously. The authors wish to advise that phone users should be safety conscious and at alert so as notice any of the factors outlined in this paper and equally apply the preventive measures as suggested.

References

Aaruni Goel, M. S. (2012). The Approaches to Prevent Cell Phone Cloning in CDMA Environment. *International Journal of Computer Applications (09 75-8887) volume 45-No. 21*, p.16.

Akash Kumar Mahato, Kumar and Akashdeep Singh (2014). Mobile Phone Cloning. *International Journal for Research in Applied Science and Engineering Technology (ijraset)*, p.224.

Eureka .S. (2017). Mobile Phone Cloning. *International Journal of Scientific & Engineering Research Volume 8, Issue 5*, p.24.

Manjula .D. and Rajanna .M. (2015). Implementing Mobile Phone Cloning in GSM and CDMA Technology. *International Journal of Innovative Research in Computer and Communication Engineering*, p.11.

Manjush Talmale, Abhishek Kinhekar, Akshay Saraf and Milind Bhajan (2013). Mobile Phone Cloning: History, Present Scenario and Precautionary

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Techniques. *International Journal of Application or Innovation in Engineering & Management (IJAIEM) ISSN 2319* , p.4.

Mishra Sandip and Nilesh K. Modi (2014). False Base Station Attack in GSM Network Environment . *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, p.3846.

Nidhi Tanwar and Sachin Chauhan (2015). Mobile Phone Cloning. *International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438*, p.350.

Mislan, R., Casey, E., & Kessler, G. (2010). The Growing Need for on - Scene

Triage of Mobile Devices. *Journal of Digital Investigation*, p.6.

Time Efficient Data Migration among Clouds

Syeda Munazza Marium, Liaquat Ali Thebo, Syed Naveed Ahmed jaffari
Computer System Engineering Department
Mehran University of Engineering & Technology
Sindh Pakistan
munazza.syed_1@yahoo.com,
liaquat.thebo@faculty.muet.edu.pk,
Naveed.jaffari@faculty.muet.edu.pk

Muhammad Hunain Memon
School of Information Science and Technology
University of Science and Technology of China,
Hefei, China
hunainmemon@ieee.org

***Abstract*— Cloud computing is one of the chief requirements of modern IT trade. Today's cloud industry progressively dependent on it, which lead mutually abundant solutions and challenges. Among the numerous challenges of Cloud computing, cloud migration is one of the major concern, and it is necessity to design optimize solutions to advance it with time. Data migration researchers attempt to move data concerning varying geographical locations, which contain huge data volumes, compact time limit and problematical architectures. Researchers aim to transfer data with minimal transmission cost and used various efficient scheduling methods and other techniques to achieve this objective. In former research struggles, numerous solutions have proposed. In our proposed work, we have explore the contextual factors to accomplish shorter transmission time. Entity Framework Core technology is utilize for conceptual modelling, mapping and storage modelling. Meant for minimum transmission cost Object Related Mapping is designated. Desired objective to achieve time efficiency during data migration has been accomplished. Results obtained when data transmission occur among azure and gearhost cloud with implementation of proposed framework with some size limitations.**

Keywords-component; Cloud Computing, Cloud Migration, Entity Framework Core, Object Related Mapping, Structure Query Language, Data Migration

I. INTRODUCTION

This Cloud computing emerging very speedily as a pervasive computational paradigm. It is a software package that continuously gaining recognition and acceptance as a resource of economical and dependable computing solution and services through internet. According to the statistical analysis worth of cloud computing market is billion dollars. It has three models which are IaaS, PaaS, SaaS (Infrastructure as a service), (platform as a service) and (software as a service) respectively. Whereas Iaas provide (storage, network, CPU, etc.) as a service.

Nowadays, a large number of applications store their data on storage servers, these applications include sensor networks, search engine cluster, video on demand servers and grid computing. In such application data migration among clouds demands lots of services, because every server run on different protocol.

In this paper they focused on an existing cloud computing technology and also forthcoming inquiries in this area and in different cloud environment, explore, service equivalence [35].This researched focused on an issue called cloud migration. Transferring data from one cloud to another with efficiency and operational processing is aimed here. Proposed online lazy migration (OLM) algorithm and a randomized fixed horizon control (RFHC) algorithm as a solution for cost-minimization problem in data migration [31]. This work is dedicated to key challenges emerged when dealing with Iaas Infrastructure as a service and networking architecture of cloud like Software-defined networking (SDN) and other architectures [29]. In this paper author investigates mobile cloud architecture and present critical analysis over application model classification, decision making entities, execution delay, cloud application models and mobile synchronization policies [32]. Author addresses the security issues when we upload data on cloud and migrate it (i-e privacy-preservability, accountability, Integrity, confidentiality, availability) [30]. This paper covers energy efficiency domain of cloud computing separated into two domains Server and network. It determine and show correlation among various domains of ICT related to energy efficiency [24]. Here author puts light on the cloud interoperability matter, discuss band of challenges like resource availability and scalability, avoiding vendor lock, interoperability, low latency and other legal issues [27]. This paper is a review of Cloud Computing (CC) and Information Technology Outsourcing (ITO) address variance among infrastructure and software services, utilization of cloud self-service and emerging role of IT in it and coded contributing elements which effect these decision [23]. Following paper author discourse about an important issue, the heterogeneous mobile platform for cloud computing. Performed examination on origins of Mobile Cloud Computing (MCC) heterogeneity factors like vendors, platform, network API and other features and recognized many challenges, to overcome these limitations analyses different architectures SOA, virtualization and middleware, etc. [25]. Author considers both spectrum efficiency and pricing efficiency of cloud and analyse power and interference management in cloud network. Proposed iterative algorithm as a solution to achieve steadiness [21]. This survey focused resource scheduling architecture of cloud computing. The survey classified resource scheduling in three categories (a) application layer scheduling (b) virtualization layer scheduling (c) deployment layer scheduling [19]. Increasing traffic requirements decrease energy efficiency of a cloud. To increase

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

energy efficiency three approaches are presented (a) MIMO (b) Dynamic spectrum access technology (c) Design frequency reuse scenario by creating smaller cells. Cloud Radio Access Network (C-RAN) with multimode support is a new model to achieve efficiency.[17]Computation offloading classify into three manners .(a) remote cloud service (b) opportunistic ad hoc cloud service and (c) connected ad hoc cloud service [18]. Virtual migration scenario is demonstrated for private cloud of organization whose applications reside on network .VM produce efficiency by removing data duplication during data migration in active and in active state [15]. Cloud resource management and scheduling performance achieve by implementing QoS-constrained algorithm on static and dynamic load which improve resource utilization and security [16]. Analyseproposed big data migration framework in current scenario using APACHE, HADOOP and SPARK utilizing various data processing schemes and machine learning algorithms, and present 5G wireless architecture prototype to process huge amount of data[9] .Flexibility of CC achieve by infrastructure virtualization. Mainly it depends on cloud computing infrastructure and design and development of application [13].

II. CLOUD COMPUTING

Earlier Network Diagrams used cloud as a symbol to represent Wide Area Network (WAN) with this context word cloud used with the internet. Now term cloud and computing used together, although cloud represent internet and computing embody services .It is blend of several amenities, which involve application development platform, shared pool resources, system management, Scalability, and multiple other services enlisted in Figure 1.

A. Key Advantages of Cloud Computing

Cloud computing delivers computational services like software integration, server space, database, storage space, backup, recovery, analytics and etc. These services are paid according to the demand and usage.

- Budget Effective: CC is cost proficient. There is no need to buy costly hardware and software. You can use them on cloud, according to Pay-as-you-go Plans. It saves money and provide well-organized resource utilization as per need.
- Universal scale: Cloud data centers are universal they are spread globally so accurate amount of bandwidth, resources can be delivered at any time. From any geographic area, one can access cloud resources.
- Throughput: Maximum output can be achieved because IT team has no burden to manage hardware, software integrity and resource handling. So entire time can be spent to achieve business goals.
- Speed: It is so quick to get access of any hardware and software service according to demand on cloud in just few minutes. So it makes so flexible to acquire such resources.
- Performance: The cloud data centers equipped with the latest technology hardware and software resources and always up to date. So these resources deliver high performance computing for your business goals.
- Trustworthiness: CC comes with backup, data recovery and disaster plan which introduce more reliability and less risk factor in computing when we are using resources from the cloud. Cloud provider network mirrored our data in a secure environment.

Figure 1. Cloud Computing Advantages.

B. Categories of Cloud Services

CC services mainly divided into three classes. Every so often is known as computing Stack because they are in stack manner. They are dependent, but different from each other explained below and shown in Figure 2.

- Infrastructure-as-a-service (IaaS): Provide virtualized computing resources like network bandwidth, processor cycles and host virtualized infrastructure using hypervisor. This class can rent infrastructure, network, server, storage, virtual machines etc. from a cloud as according to your demand. IaaS vendors include Amazon, Rackspace, Cloud Foundry.

Identify applicable sponsor/s here. *(sponsors)*

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Platform as a service (PaaS): It deliver adaptable and optimal environment. For development purpose this service is used to develop applications on cloud platforms so there is no need to worry about databases, storage, network and other resources essential for software development some examples of PaaS vendors include Microsoft Azure, Amazon, Force.com.
- Software as a service (SaaS): This distribution model is persistent, accessible, and scalable eliminate expense of Licensing, maintenance provisioning etc. It allow to use desired application over the cloud. Application maintenance, security and updates is not consumer nuisance it is a responsibility of cloud service providers. Consumer only need browser to connect with it. There are many examples of SaaS vendors – Salesforce.com, Google Apps, Ning, Cenzic and etc.

Figure 2. Cloud Computing Services

C. Equations

CC services can be deployed in three different ways.

- Public cloud: In public cloud after deployment of your data you are not responsible to manage, maintain and host your data in the datacentre.
- Private Cloud: These clouds are also called as enterprise clouds, it is managed and maintain via internal sources in a private environment.
- Hybrid Cloud): It is a combination of public and private clouds. Data can be moved among both clouds with permission it provides flexible environment.

III. Importance of Cloud Migration

In every organization database technology is fundamental tool. The need and importance of databases are growing with the growth of IT technology. Powerful database systems have been introduced. As the world is becoming a global village, many organizations are handling their business from remote locations. As this is an era of cloud computing organization keep their data in the cloud, moreover there is a need arise to migrate data among the clouds. Whereas Data migration is a process where we extract, clean, transform and upload data into the new platform [8].

A. Cloud Migration Process:

These few steps will be followed by cloud migration process shown in Figure 3:

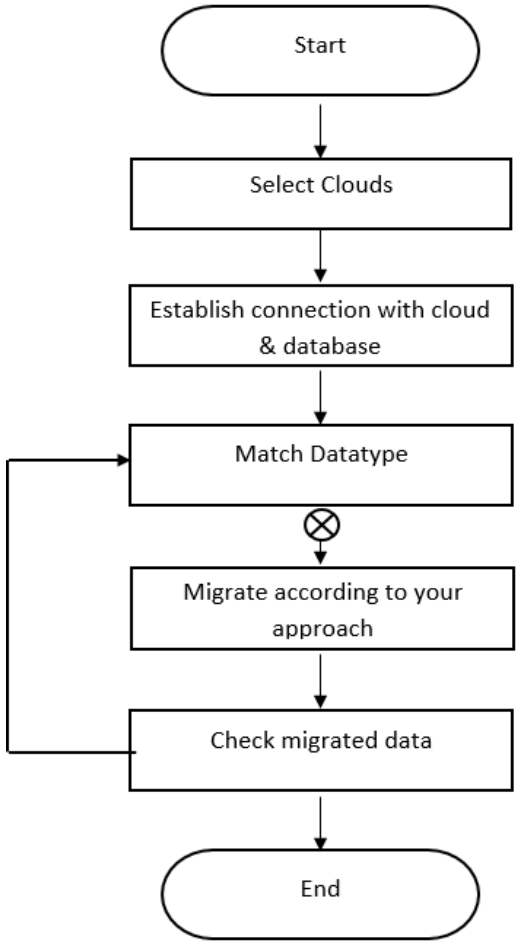


Figure 3. Cloud Migration Procedure [38]

Before migration below mention parameters of the database must be examined.

- Format Check: Check level of consistency and usability of the database.
- Consistent data: Exclude data which disrupt the logical consistency during the commit procedures.
- Length Check: examine length of your data when converting from one type to another.
- Range Check: Scrutinize the length of your data to assign an appropriate data type for memory efficiency.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Integrity Check: Check integrities for the data association scheme.

B. Data Migration Approaches

Previously, many approaches used for migration of data between systems. Many tools are fabricated with DBMS to accomplish this. Migration categorizes in three classes:

(a) By tools
(b) Manual data Migration
(c) User define new System for data migration and etc

According to research published in 2017 following technologies used by the developers cloud migration [7]. Mention in TABLE I.

TABLE I
TECHNOLOGIES USED FOR CLOUD MIGRATION

1	Manual Coding	174/**64%**
2	Data Integration Tool	145/**54%**
3	Data Mapping Tool	96/**36%**
4	Data Dictionary	73/**27%**
5	Testing Tool	52/**19%**
6	Data Archiving Tool	29/**11%**
7	Other	15/**16%**

In our research, we want transfer data among clouds so we will use a third approach means we will program our user defined system to transfer data among the clouds.

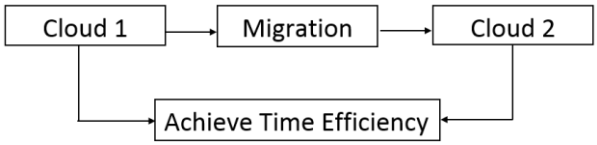


Figure 4. Cloud Migration Flow

IV. PROBLEM STATEMENT

When there is data transfer between cloud rate of accuracy and speed is a main challenge .Previous researches transfer table data in columns and Obtained time efficiency [38]. In this research text files as well as image files would be transferred among cloud to accomplish time efficiency. Flow of migration strategy shown in Fig. 4.

A. Data Integration:

Data integration is a scenario where data from various sources combine at one place, and produce a unified and meaningful view of data. Design such a system with maximum efficiency is a challenge. It is usually characterized by an architecture.

(a) Architecture of global schema
(b) Architecture of set of sources

Global schema contains a virtual view of data while the sources contain a the real data [42].The difficult part in designing a data integration system is to design source depiction writing, mapping and schema which require expertise knowledge to write scripts [34].

Factors for Data Integration:

(a) XML
(b) Elastic Query Processing
(c) Model Organization
(d) Data organization
(e) P2P Data Organization

- XML (extensible mark-up language): XML is one of the pillars in the development of data integration. Reason behind this is its syntactic format so data can easily share among sources. For good data integration the system should capable of1) XML (extensible mark-up language): XML is one of the pillars in the development of data integration. Reason behind this is its syntactic format so data can easily share among sources. For good data integration the system should capable of handling complex XML. So it was challenging to design such a system which can interpret Nested and complex HTML tags [40]
- Elastic query processing: When a query posted on a schema and its task is to gather data from a set of sources to create a view. In order to form that view the query must be very efficient. There are many techniques available for efficient query processing, but in data integration system these techniques not completely applicable because its optimizer not contained detailed information so good path selection is not possible sometimes for query execution.
- Model Organization: For management of data integration system main task is to design a mapping route among schemas. So algebraic operation would be a solution to map between schemas, but it is a complex task.
- P2P Data Organization: Peer to peer is an architecture which shares data by means of distributed contrivance. In a distributed system complex semantic mappings are programmed individually among a set of sources and global schema network paths.

B. Data Access Technique

These are data mapping techniques which used to access data from one or more sources. The data can be accessed from tables in a different manner. Here we will discuss about few data mapping techniques to access data

(a) Semantic Mapping
(b) Data Driven Mapping
(c) Object Related Mapping

- Semantic Mapping: This technique detects and discover exact matches. Any logical alteration of column data cannot be handled and recognized by it. Semantic system signifies information associated with a particular domain and concepts. It can provide, capable and

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

potential domain dependent system schema to access data [10]. This is a process where an object, action, identities devoted to entities and algebraic sequences mapped them shown in Figure 5.The disadvantage of this technique is that it is domain dependent, does not support logic implementation of data and do not consult the metadata registry for synonym search.

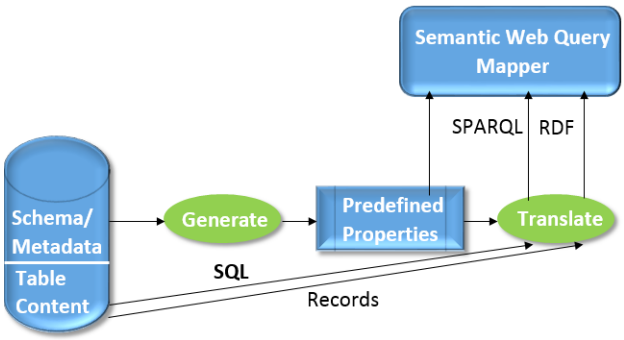


Figure 5. Semantic Mapping [43]

- Data-Driven Mapping: This technique evaluates data two times. (a) Heuristic evaluation (b) Statistical evaluation, so it is capable of automatic discovery complex mapping routes among a set of sources for data access and data integration [10]. With this approach set of sources can produce results which contain concatenations, arithmetic calculations, logical transformations, substrings, other data manipulation operations as shown in Figure 6. It can also define and handle exceptions.

Figure 6. Data Driven Mapping [45]

- ORM Mapping: While need arises to transfer data between systems which are not compatible with each other, then the technique Object Relational Mapping (ORM) is used. Object oriented programming is used to create maps with relational database systems, XML repositories and other data sources, etc. Virtual object databases, create with this technique so data can be mapped and access [10]. Mostly database system deals with scalar values for extract, transform, load (ETL) and data manipulation operations so it is necessary that the object must behave both as an object and scalar value as per as need. As we have a logical representation of the object here so difficult part is that to store them in the database as an object and it should be capable of preserving their properties and its relationship with entities for reuse as scenario described in Figure 7(a) and Figure 7(b).

Figure 7(a). Object Related Mapping

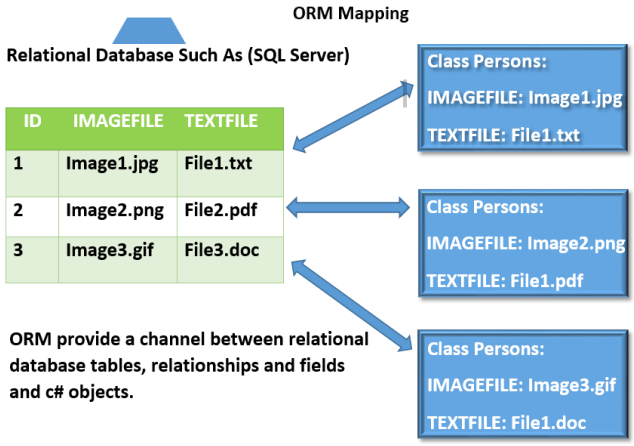


Figure 7(b). Object Related Mapping

V. METHODOLOGY

For migration of data from azure cloud to gearhost cloud so we are using ORM technique. We have interact with object for efficient data migration. We will follow the entity data model and entity frame core (EF core) technology as shown in Fig. 9 for our research work.

A. Entity Data Model

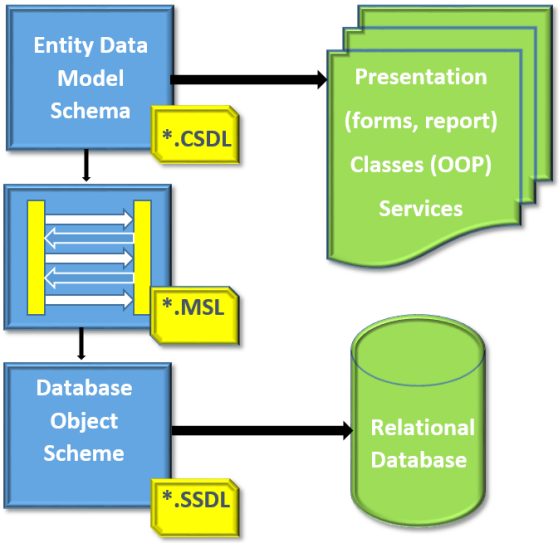
This model do not consider in which form data is stored it focuses on a concept regarding the structure of data.

- Conceptual Schema: This defines the scenario where we examine our entities, and decide which entity classes are required to accomplish tasks and establish relationship among them. By following these steps here we will produce a high level view of our database.
- Mapping: Both designed conceptual and storage schema will be mapped at this stage.

This model describes a relational model and data storage representation in Figure 8. Processing performed using Conceptual schema definition language (CSDL), mapping specification language (MSL), and store schema definition language (SSDL).

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 8 .Entity Relationship Diagram [44]



A. *EDM Data Structure Concepts*

This model do not consider in which form data is stored it focuses on a concept regarding the structure of data.

- Entity: Structure of data describes in EDM come under the umbrella of this concept.
- Association: Referential concepts in tables describe through different association type in our database. Like Primary key, Foreign key, composite key, etc.
- Property type: Here we define property to the pre-defined entities. Like here we will define the data type of the columns, which can be primitive or complex.

B. *Data Loading*

For loading data into application from your database we have three types of loadings.

(a) Lazy Loading
(b) Explicit Loading

(c) Eager Loading.

- Lazy Loading: This is default data loading. It loads the main entity reference in the query instead of related entity reference in the query, which makes it slow.
- Explicit Loading: If we disable lazy loading so we can manually add related entity references to the query using load () method.
- Eager Loading: It loads all the related entities. But after the loading of the main entity of query. For this purpose, its use includes () and then include () methods. Advantage of eager loading is that it extracts massive data in one query processing.

VI. EF CORE(ENTITY FRAMECORE WORK)

It lightweight, extensible technology and it support cross platform development. It can also provide functionality of ORM mapper by using objects created in .net application code. It supports many database engines like SQL Server, SQLite, Oracle, Microsoft Access files, PostgreSQL and etc. In our research we are dealing with SQL Server. [34]

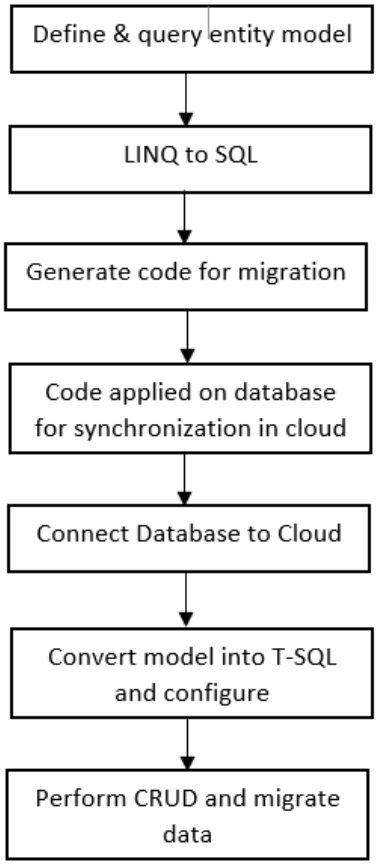


Figure 9 .Methodology

Steps Shown in Figure 9 are the operations we have perform to achieve cloud database migration.

A. *Prerequisite*

Installed following application packages before starting configuration and ORM class and object defining scenario. Microsoft.EntityFrameWorkcore.SqlServer Microsoft.EntityFrameWorkcore.Tools -Pre Microsoft.EntityFrameWorkcore.SqlServer.Domain [34].

B. *Creating Model*

We have Fluent API configuration for creating the entity model. The Model builder API used to configure it. For defining properties data annotations have been used. Entity model has metadata of type blog.

C. *Keys Constraint*

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Primary keys and other unique identifiers are defined using the model builder API without effecting entity classes.

D. *Value Generation*

Value generated by the user in the database every time we populate database entities with data and that value is saved using savechanges () method.

E. *Required Entities*

In our framework 7 entities have been defined PersonalID, LastName, FirstName, Address, City, TextFile, Picture.Among them only required entity is PersonalID entity .Remaining entities left optional in order to test result in three different categories (a) column text data transfer (b) mage file migration (c) text file migration.

F. *Relational Modelling*

: For relational database modeling of discussed framework this package (Microsoft.EntityFrameworkCore.Relational package) is installed.

G. *Table Mapping*

The above mentioned package is used for table mapping DbSet<TEntity>. If table entity is not defined in dbset context, then class name will use for mapping [34].

H. *Column Mapping*

This mapping is performed to map column, that after query which column data should fetch and save in table of other database after migration.

I. *Data Mapping*

Datatype of columns are also mapped among databases as we are using the same database in different clouds so we will map datatype and their maximum length. Here we are using dbo schema in our databases.

VII. Migration

We have used Migration builder API for data transfer by using Migration Builder Operations () following up method approach.

VIII. Results

In our research, we have migrated data from one cloud to another both having different architectures. We are using EFcore technology for this purpose and ORM technique for object mapping. We will observe time efficiency in results when we compare it with and without EFcore and ORM.

TABLE II
FORMULAE

Formulae For Result Calculations
Save Time Efficiency =Save Old-Save ORM
Transfer Time Efficiency =Transfer Old-Transfer ORM
Total Time Efficiency =Save Time Efficiency+Transfer Time Efficiency

Whereas stopwatch() method is used to calculate time consumed by query execution in milliseconds.

TABLE III
IMAGE GB TIME EFFICIENCY ANALYSIS

ISTE(GB)	**ITTE(GB)**	**ITTTE(GB)**
6478 ms	13048 ms	19526 ms
-994 ms	1557 ms	563 ms
5236 ms	-1768 ms	3468 ms
662 ms	18145 ms	18807 ms
-269 ms	11171 ms	3153 ms
-1329 ms	8838 ms	-2435 ms
4072 ms	-4589 ms	-517 ms
295ms	-3949 ms	-3654 ms
1111 ms	-1315 ms	-204 ms
854ms	896ms	1750ms

ISTE=Image Save Time Efficiency, ITTE= Image Transfer Time Efficiency, ITTTE= Image Total Transfer Time Efficiency, GB=gigabytes, ms=milliseconds

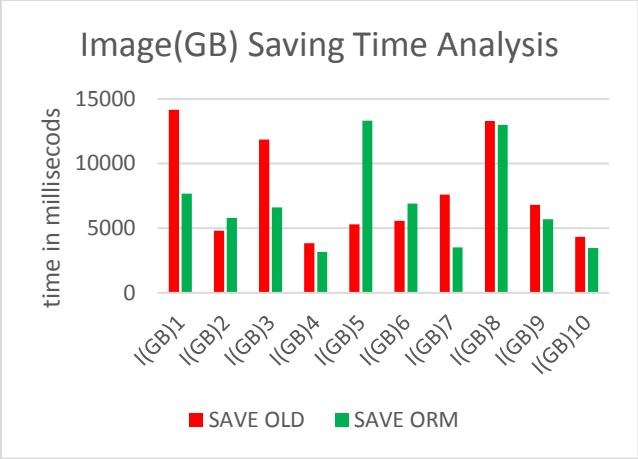


Figure 10. Image (GB) Saving Time Analysis, I(GB)=Image in GB

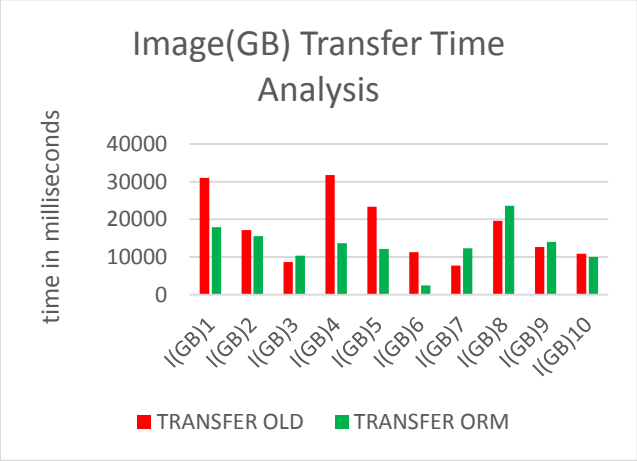


Figure 11. Image (GB) Transfer Time Analysis

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

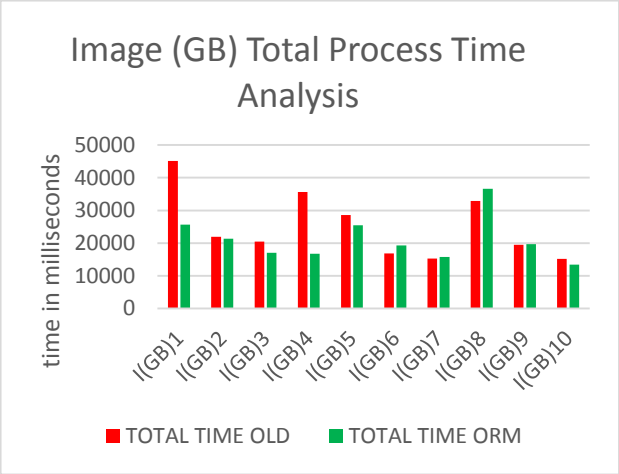


Figure 12. Image (GB) Total Process Time Analysis

TABLE IV
IMAGE KB TIME EFFICIENCY ANALYSIS

ISTE(KB)	ITTE(KB)	ITTTE(KB)
-74 ms	-508 ms	-582 ms
2516 ms	1511 ms	2326 ms
-2126 ms	1114 ms	-1012 ms
1364 ms	8603 ms	9967 ms
375 ms	120 ms	495 ms
577 ms	3649 ms	4226 ms
1077 ms	5921 ms	6998 ms
4735 ms	2228 ms	6963 ms
1059 ms	3159 ms	4218 ms
412 ms	1212 ms	1625 ms

ISTE=Image Save Time Efficiency, ITTE= Image Transfer Time Efficiency, ITTTE= Image Total Transfer Time Efficiency, KB=kilobytes, ms=milliseconds

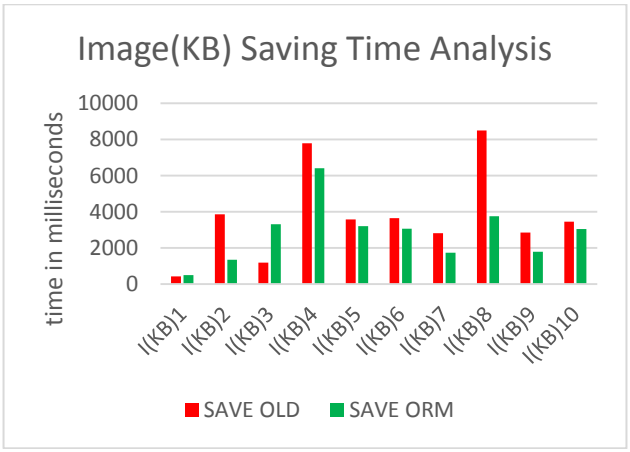


Figure 13. File (KB) Saving Time Analysis, I(kB)=Image in KB

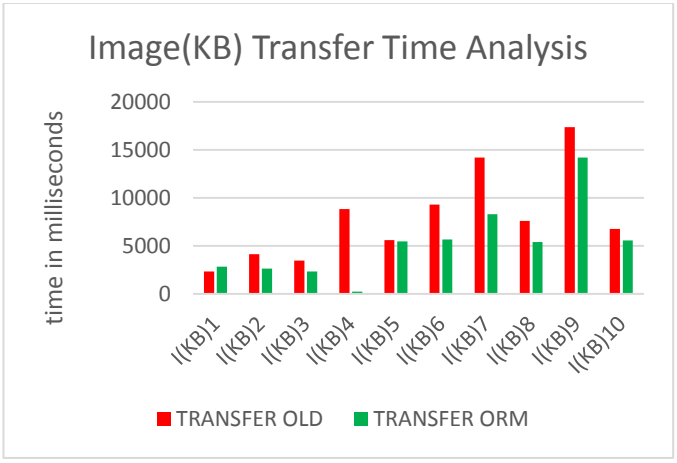


Figure 14. File (KB) Transfer Time Analysis

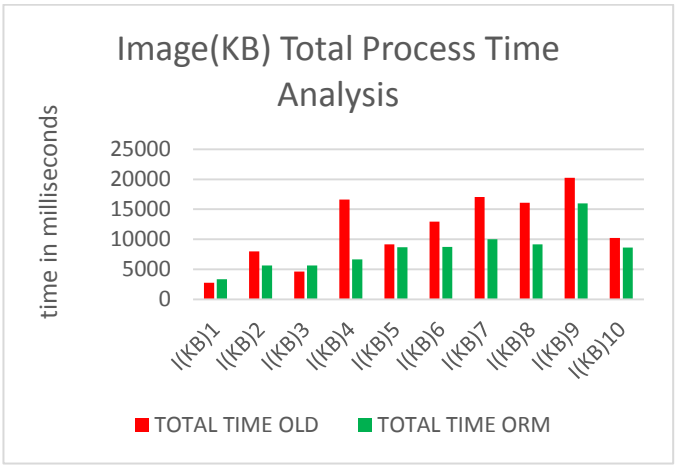


Figure 15. File (KB) Total Process Time Analysis

TABLE V
FILE KB TIME EFFICIENCY ANALYSIS

FSTE(kb)	FTTE(kb)	FTTTE(kb)
-167ms	427 ms	260 ms
936 ms	-516 ms	420 ms
333 ms	829 ms	1162 ms
2871 ms	975 ms	3846 ms
3309 ms	-560 ms	2749 ms
672 ms	-236 ms	436 ms
1816 ms	222 ms	2038 ms
-207 ms	-996 ms	-1203 ms
1446 ms	-446 ms	1000 ms
81 ms	140 ms	221 ms

FSTE=File Save Time Efficiency, FTTE= File Transfer Time Efficiency, FTTTE= File Total Transfer Time Efficiency, KB=kilobytes, ms=milliseconds

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

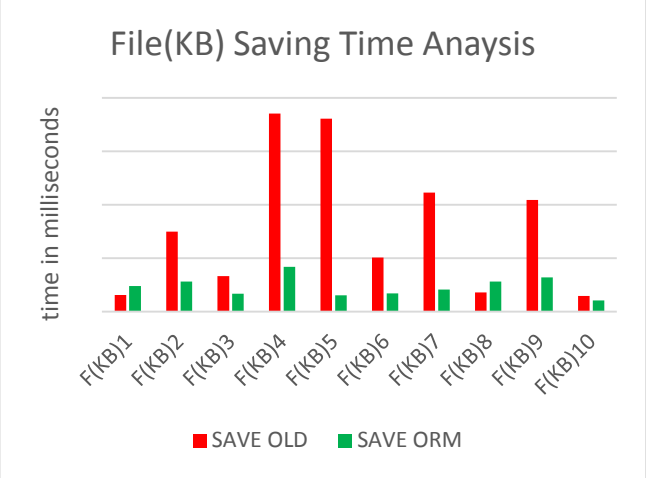


Figure 16. File (KB) Saving Time Analysis, F(MB)=File in KB

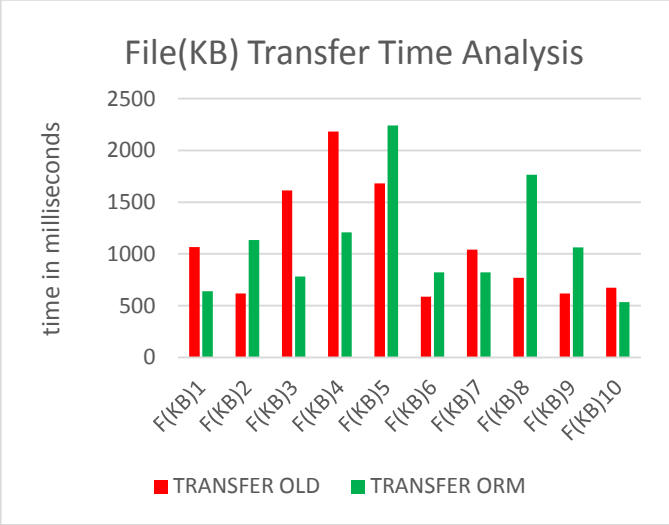


Figure 17. File (KB) Transfer Time Analysis

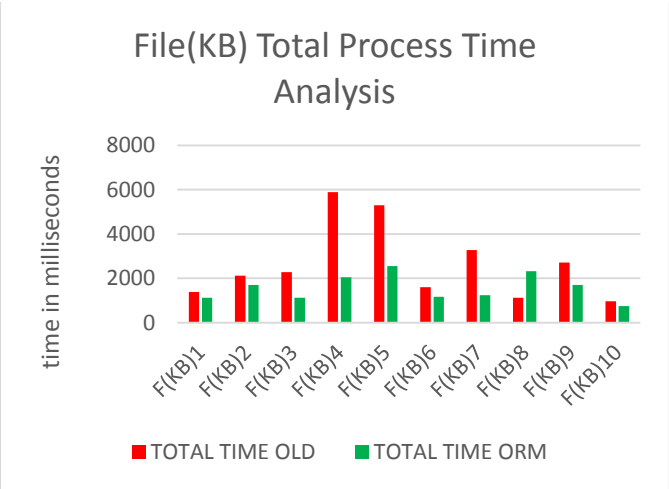


Figure 18. File (KB) Total Process Time Analysis

TABLE VI
FILE GB TIME EFFICIENCY ANALYSIS

FSTE(GB)	FTTE(GB)	FTTTE(GB)
-1089 ms	131 ms	1473 ms
23 ms	1126 ms	3142 ms
9680 ms	198 ms	10696 ms
-33 ms	121 ms	1030 ms
10439ms	-608 ms	11605 ms
37 ms	12 ms	938 ms
9901 ms	223 ms	10936 ms
38 ms	321 ms	1260 ms
-78 ms	420 ms	1384 ms
151 ms	-225 ms	74 ms

FSTE=File Save Time Efficiency, FTTE= File Transfer Time Efficiency, FTTTE= File Total Transfer Time Efficiency, GB=gigabytes, ms=milliseconds

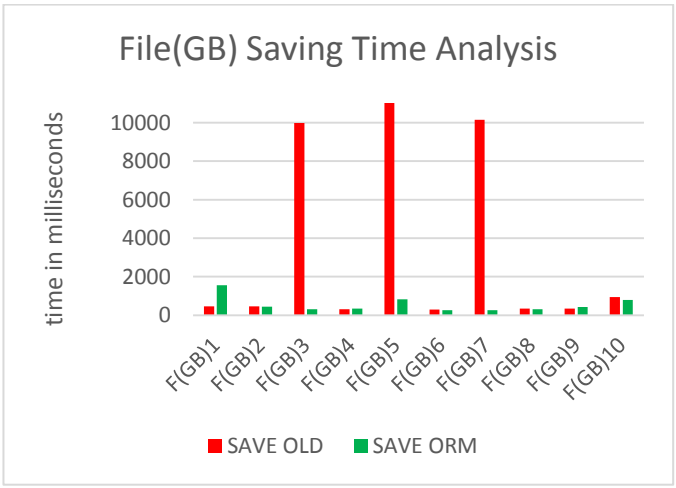


Figure 19. File (GB) Saving Time Analysis, (GB)=File in GB

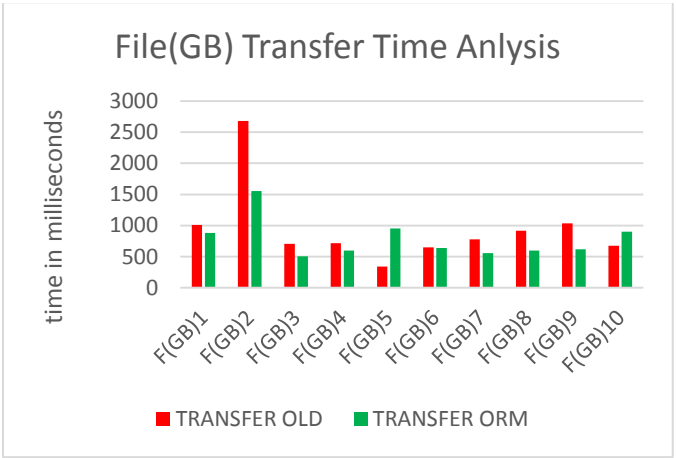


Figure 20. File (GB) Transfer Time Analysis

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

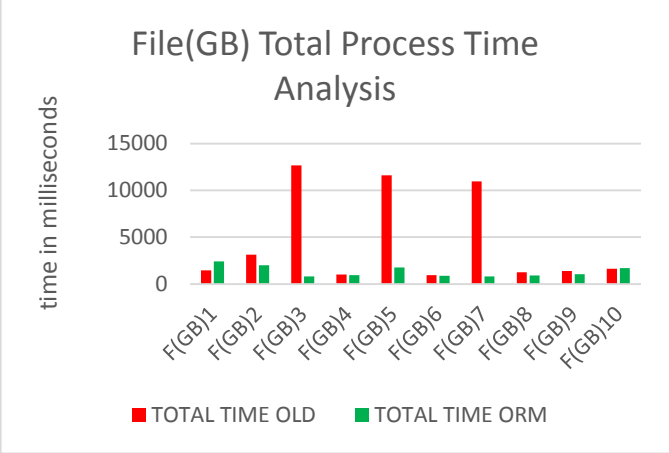


Figure 21. File (GB) Total Process Time Analysis

Results calculated in four categories (a) images in MB (b) images in KB (c) Text Files in MB (d) Text Files in KB. We have tested each category with 100 input files. But here we are showing results of only 10 individual cases in each category to demonstrate efficiency because it is not possible to show the results of all input files also taking the average of these input files can mislead to a conclusion.

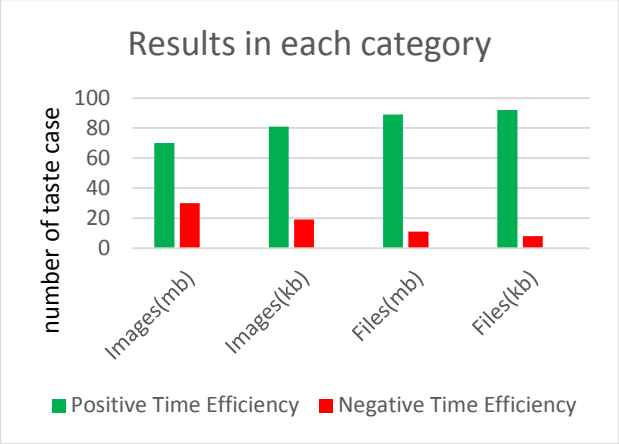


Figure 22. Aggregate Result Analysis

The above chart shows the number of successful and failed Cases in each category.

IX. Discussion

As we have mentioned values in tables from TABLE II to TABLE V and graphs can be observed from Fig. 10 to Fig. 21 in most of the cases ORM method is efficient then old method but in some cases very attention-grabbing results are found when difference in their time is more than 200% or more than that as we can see in Fig. 10. old method is taking very much less time in migration than ORM also investigating Fig. 19 found that old method is consuming very high time where as new method consumed few milliseconds for it. But interesting thing is that such cases are found when we are dealing with Images in GB and Images in KB.

X. Conclusion

From the above results multiple things can be observed like sometimes Save Time Efficiency is negative and Sometimes Transfer Time Efficiency is negative so both of these factors are responsible for negative results. Whereas, one more important factor is speed of internet if good internet speed is constant, results can be more constant. Also image migration consuming more time than text file migration. If we combine the results of these 4 categories statistically we can say that we have achieved approximately 80% efficiency.

XI. Future Work

It is still an ongoing research area in future instead of two clouds, multiple clouds can be used and test results, to find out the appropriate reason of negative efficiency. As we have transfer input from 100kb to 8GB. One can target to transfer images and files more than 8 GB to achieve positive time efficiency.

Acknowledgment

I would like to thank Muhammad Bilal Amjad from Microsoft for providing research resources of azure cloud.

References

[1] A. M, S. W, and K. K. L, “L iveS ervice Migration in Mobile Edge C louds,” pp. 2–9, 2017.
[2] G. J. L. Paulraj, S. A. J. Francis, J. D. Peter, and I. J. Jebadurai, “Resource-aware virtual machine migration in IoT cloud,” Futur. Gener. Comput. Syst., vol. 85, pp. 173–183, 2018.
[3] M. C. Silva Filho, C. C. Monteiro, P. R. M. Inácio, and M. M. Freire, “Approaches for optimizing virtual machine placement and migration in cloud environments: A survey,” J. Parallel Distrib. Comput., vol. 111, pp. 222–250, 2018.
[4] S. Zimányi, E., Jallow, B., Kashef, “Object Relational Mapping and Entity Framework,” pp. 1–16, 2018.
[5] N. Sunderhauf, T. T. Pham, Y. Latif, M. Milford, and I. Reid, “Meaningful maps with object-oriented semantic mapping,” IEEE Int. Conf. Intell. Robot. Syst., vol. 2017–September, pp. 5079–5085, 2017.
[6] H. reza Bazi, A. Hassanzadeh, and A. Moeini, “A comprehensive framework for cloud computing migration using Meta-synthesis approach,” *J. Syst. Softw.*, vol. 128, pp. 87–105, 2017.
[7] D. M. Pro, “Data Migration Research Study,” 2017.
[8] A. Botta, W. De Donato, V. Persico, and A. Pescapé, “Integration of Cloud computing and Internet of Things: A survey,” *Futur. Gener. Comput. Syst.*, vol. 56, pp. 684–700, 2016.
[9] E. Zeydan et al., “Big data caching for networking: Moving from cloud to edge,” IEEE Commun. Mag., vol. 54, no. 9, pp. 36–42, 2016.
[10] A. G. Huth, W. A. De Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant, “Natural speech reveals the semantic maps that tile human cerebral cortex,” Nature, vol. 532, no. 7600, pp. 453–458, 2016.
[11] Q. Yan, F. R. Yu, S. Member, Q. Gong, and J. Li, “Software-Defined Networking (SDN) and Distributed Denial of Service (DDoS) Attacks in Cloud Computing Environments : A Survey , Some Research Issues , and Challenges,” no. c, pp. 1–23, 2015.
[12] B. C. Stahl, J. O. B. Timmermans, and B. D. Mittelstadt, “The Ethics of Computing : A Survey of the Computing-Oriented,” vol. 48, no. 4, 2016.
[13] C. Colman-meixner, C. Develder, M. Tornatore, and B. Mukherjee, “A Survey on Resiliency Techniques in Cloud Computing Infrastructures and Applications,” no. i, 2016.
[14] [14] Y. Li, K. Gai, L. Qiu, M. Qiu, and H. Zhao, “Intelligent cryptography approach for secure distributed big data storage in cloud computing,” *Inf. Sci. (Ny).*, vol. 387, pp. 103–115, 2017.
[15] N. Bila *et al.*, “Energy-Oriented Partial Desktop Virtual Machine Migration,” *ACM Trans. Comput. Syst.*, vol. 33, no. 1, pp. 1–51, 2015.

[16] X. Wu, G. F. Liu, J. J. Xu, and Ieee, "A QoS-Constrained Scheduling for Access Requests in Cloud Storage," *Proc. 2015 10th Ieee Conf. Ind. Electron. Appl.*, pp. 155–160, 2015.
[17] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): A primer," *IEEE Netw.*, vol. 29, no. 1, pp. 35–41, 2015.
[18] M. Chen, Y. Hao, Y. Li, C. Lai, and D. Wu, "O N THE C OMPUTATION O FFLOADING AT A D H OC C LOUDLET :," no. June, pp. 18–24, 2015.
[19] Z.-H. Zhan, X.-F. Liu, Y.-J. Gong, J. Zhang, H. S.-H. Chung, and Y. Li, "Cloud Computing Resource Scheduling and a Survey of Its Evolutionary Approaches," *ACM Comput. Surv.*, vol. 47, no. 4, pp. 1–33, 2015.
[20] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Rob. Auton. Syst.*, vol. 66, pp. 86–103, 2015.
[21] Z. Yin, F. R. Yu, S. Bu, and Z. Han, "Joint cloud and wireless networks operations in mobile cloud computing environments with telecom operator cloud," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 7, pp. 4020–4033, 2015.
[22] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges," *Inf. Sci. (Ny).*, vol. 305, pp. 357–383, 2015.
[23] S. Schneider and A. Sunyaev, "Determinant factors of cloud-sourcing decisions: Reflecting on the IT outsourcing literature in the era of cloud computing," *J. Inf. Technol.*, vol. 31, no. 1, pp. 1–31, 2016.
[24] T. MASTELIC, A. OLEKSIAK, H. CLAUSSEN, I. BRANDIC, J.-M. PIERSON, and A. V VASILAKOS, "Cloud Computing: Survey on Energy Efficiency.," *ACM Comput. Surv.*, vol. 47, no. 2, p. 33:1-33:36, 2015.
[25] S. A. Z. Sanaei A. Gani, R. Buyya, "Heterogeneity in Mobile Cloud Computing:Taxonomy and Open Challenges," *IEEE Commun. Surv. Tutorials (Accepted Publ.*, vol. 16, no. 1, pp. 1–24, 2013.
[26] L. Wei *et al.*, "Security and privacy for storage and computation in cloud computing," *Inf. Sci. (Ny).*, vol. 258, no. May, pp. 371–386, 2014.
[27] A. N. Toosi, R. N. Calheiros, and R. Buyya, "Interconnected Cloud Computing Environments," *ACM Comput. Surv.*, vol. 47, no. 1, pp. 1–47, 2014.
[28] P. Pant and S. Thakur, "Data Migration Across The Clouds," *Int. J. Soft Comput. Eng.*, vol. 3, no. 2, pp. 14–21, 2013.
[29] S. Azodolmolky, P. Wieder, R. Yahyapour, and G. Wissenschaftliche, "Cloud Computing Networking : Challenges and Opportunities for Innovations," no. July, pp. 54–62, 2013.
[30] Z. Xiao, Y. Xiao, and S. Member, "Security and Privacy in Cloud Computing," vol. 15, no. 2, pp. 843–859, 2013.
[31] L. Zhang, C. Wu, Z. Li, C. Guo, M. Chen, and F. C. M. Lau, "Moving Big Data to The Cloud : An Online Cost-Minimizing Approach," vol. 31, no. 12, pp. 2710–2721, 2013.
[32] R. Khan, M. Othman, S. A. Madani, and I. Member, "A Survey of Mobile Cloud Computing Application Models," pp. 1–21, 2013.
[33] B. Di Martino and A. Esposito, "Semantic and Agnostic Representation of Cloud Patterns for Cloud Interoperability and Portability," 2013.
[34] N. G. Fielding, "Triangulation and Mixed Methods Designs : Data Integration With New," 2012.
[35] W. Venters and E. A. Whitley, "researching desires and realities," vol. 27, no. 3, pp. 179–197, 2012.
[36] S. Sakr, A. Liu, D. M. Batista, and M. Alomari, "A Survey of Large Scale Data Management Approaches in Cloud Environments," vol. 13, no. 3, pp. 311–336, 2011.
[37] C. Wang, K. Ren, W. Lou, and J. Li, "Toward Publicly Auditable Secure Cloud Data Storage Services," no. August, pp. 19–24, 2010.
[38] D. Maturana, P. Chou, M. Uenoyama, and S. Scherer, "Real-Time Semantic Mapping for Autonomous Off-Road Navigation," pp. 335–350.
[39] X. Luna, D. Alon, and H. Cong, *Data integration with uncertainty*. 2009.
[40] [1] A. Halevy and J. Ordille, "Data Integration : The Teenage Years," pp. 9–16.
[41] A. Doan and A. Y. Halevy, "Research in the A Brief Survey," vol. 26, no. 1, pp. 83–94, 2005.
[42] M. Lenzerini, L. Sapienza, V. Salaria, and I.- Roma, "Data Integration : A Theoretical Perspective," pp. 233–246, 2002
[43] J. F. Sequeda, S. H. Tirmizi, O. Corcho, and D. P. Miranker, "Survey of directly mapping SQL databases to the Semantic Web," *Knowledge Engineering Review*. 2011.
[44] Codemag.com. (2018). *Introducing ADO.NET Entity Framework*. [online] Available at: https://www.codemag.com/article/0711051/Introducing-ADO.NET-Entity-Framework [Accessed 14 Aug. 2018].
[45] Inc., A. (2018). *Why Data Visualization with Tableau ? - Advanz101*. [online] Advanz101. Available at: http://www.advanz101.com/data-visualization-services-with-tableau/ [Accessed 14 Aug. 2018].

Recognizing and Classifying Object colour with 16Million Colour Model Automatically

Nancy Chinyere Woods
Department of Computer Science
Faculty of Science, University of Ibadan,
Oyo Road, Ibadan, Nigeria
chyn_woods@yahoo.com

Charles Abiodun Robert
Department of Computer Science
Faculty of Science, University of Ibadan,
Oyo Road, Ibadan, Nigeria
abc.robert@ui.edu.ng

Abstract -**Object recognition is the art of determining the identity of an object being observed in a digital image and the algorithms used to recognise objects, often rely on matching, learning, or pattern recognition using appearance-based or feature-based techniques. In this paper, a feature-based technique was developed to align and identify the colour of any recognised object. Image dataset containing 998 digital Images were obtained from the Internet and personal collections. Seventy percent (1,365) of the images were used to train object recognition algorithm using the histograms of oriented gradients features. Thirty percent (585) of the images were used for testing. An algorithm was developed to locate a pixel within a recognised object to enable the identification and verification of the object's predominant colour based on the RGB colour model. The algorithm recognised and classified objects in test images. The predominant colour of the recognised object was identified with 99.88% accuracy and verified. The developed model adequately identified visible objects in images with their colour.**

Keywords: Colour identification, Object recognition, object colour

I. INTRODUCTION

One of the concerns in object recognition is the task of finding a given object in an image or video sequence [1]. This task may sometimes be dependent on the colour of the object. A scenario of searching for an image containing a "red ball" refers to the set of colours and shapes found in that image. The colour information of pixels play very important roles, in the identification of shapes, edges, corners or objects in any image. Colour perception is subjective because individuals identify colours differently. A typical reference of individuality in colour perception was reported in the article on 'The dress' that went viral in May 2015 [2]. According to that article, 'The Dress', represented the first time a single image was seen as completely different colours by so many different people. A majority of the viewers claimed the dress was "blue and black" in colour while others saw it as "white and gold", and yet some others saw different colours, other than blue and black, or white and gold. The field of human medicine explained that people perceive colours differently based on the individuals colour vision type as reported by [3]. So what happens when an individual with some form of colour blindness is asked to identify an object with a particular colour? That person's judgement may not be completely correct, because of the type of colour vision. The question is how do we identify colour in a distinctive way without individual biases?

II. TYPES OF COLOUR VISION

The type of colour vision an individual has was reported by [3] to affect the way he or she perceives colours. This is because humans have cones in their eyes which are the sensors responsible for colour vision [4]. According to experimental evidence, humans have between 6 to 7 million of these cones in the retina which can be divided into three principal sensing categories, called the red, green and blue cones [5], [3]. Some Individuals, due to colour vision defect caused by insensitive cones or colour blindness see colours differently. Consequently, there are eight generally accepted types of colour vision ranging from Normal vision to Colour blindness. These eight types are listed in Table 1 with approximate frequency of occurrence values [3]. The work of Romano (2008) illustrated what happens when individuals with some sort of colour blindness look at colours. People that are red-blind or green-blind will see the correct shape but cannot correctly identify its colour [6].

It can then be cumbersome for these individuals to search through an image database with the aim of the search as identifying an "object" with specific colours. Consequently, if a person with some level of colour blindness were to visually identify the colour of an object, his judgement will be subjective and defective [7] and could lead to poor search results. Therefore, this paper presents model for the automatic recognition of the predominant colour of any identified object in a digital image.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 1. Colour Vision Types with Approximate Frequency of Occurrence Values (Source [3])

Type (frequency)	Cones Affected	Effect on Vision
normal (91%)	red, green, and blue function normally	normal trichromatic vision
deuteranomalia (5%)	green cones are red-sensitive	green is perceived as black, red and yellow appear identical, turquoise as blue
protanomalia (1%)	red cones are green-sensitive	green is perceived as a shade of yellow, red as black, purple as blue
protanopia (1%)	red cones are insensitive	red is perceived as black, green and yellow as identical, purple as blue
deuteranopia (1%)	green cones are insensitive	green is perceived as black, red and yellow appear identical, turquoise as blue
tritanopia (10^{-3}%)	blue cones are insensitive	blue is perceived as black, purple and red are identical, turquoise appears green
monochromatopia (10^{-4}%)	only blue cones function normally	red, yellow, and green are perceived as black
achromatopia (10^{-5}%)	all colour cones are insensitive, only rods (night vision) function normally	day-blindness

III. OBJECT RECOGNITION

Object recognition starts with image segmentation. Segmentation is a method of grouping pixels in an image into similar classes to create a set of non-overlapping regions in the image [8]. Image segmentation also involves subdividing an image into its constituent parts, regions or objects [5], [9], and these regions or objects are usually portions of interest in the image [10]. The primary application of image segmentation is to isolate groups of neighbouring pixels with related properties so that these groups can then be analysed and classified together as 'objects'. The term 'objects' could represent anything ranging from house, cars, humans, to sky and gardens. The level of detail to which the subdivision is carried depends on the problem being solved and, segmentation ought to stop once the 'object' or region of interest has been identified [5]. For example, if the aim of a search is to identify an image that contains the sky, then once such an image has been subdivided into two regions as shown in Fig 1, there is no need to further subdivide the image, since the areas of interest have been identified [11]. Hence, image segmentation is the foundation of object recognition and computer vision [1], and the main aim of image segmentation is for the identification of 'objects' [12].

Object recognition is the task of finding a given object in an image or video sequence [1] from a set of known labels [13]. To identify a specific object, the algorithms used often rely on matching, learning, or pattern recognition using appearance-based or feature-based techniques [14]. Several approaches/algorithms are employed for extracting the features of the object in question which then leads to detecting and recognising the 'object'. These approaches can be classified as:- (1) Top-down approaches, which often include a training stage to obtain class-specific model features or to define object configurations; (2) Bottom-up approaches, that start from low-level or mid-level image features, like edges or segments and then build up hypotheses from such features, extend them by construction rules and then evaluate by certain cost functions [15]. The third category combines top-down and bottom-up methods in a bid to harness the advantages of both aspects.

Some of these approaches include; The Scale Invariant Feature Transform (SIFT) [16] that transforms an image into a large collection of local feature vectors, each of which is invariant to image scaling, and rotation, and partially invariant to illumination changes or 3D projection. Bag-of-words descriptor [17] which counts the frequency with which each visual word occurs in an image. Haar-like features used in the Viola-Jones algorithm for face detection that uses a sliding window to discover front faces. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image. The Viola-Jones algorithm relies on an ensemble of weak classifiers, each one of which captures some simple contrast-based feature in the candidate face window [18]. Histogram of Oriented Gradients [19], uses a single filter on features to represent an object category by using a sliding window approach, where a filter is applied at all positions and scales of an image. This approach is best suited for textured objects captured from a fairly consistent viewpoint [12]. Fig 2 shows the main components of a sliding window detector. According to [12], to learn from the training images, some feature representation must be selected and labelled examples (both negative and positive examples) are used to train a classifier that computes how likely it is that a given window contains the object category of interest. Therefore, given a fresh image, the features from each of its sub-windows at multiple scales are extracted, and then tested by the classifier to aid the identification of an object.

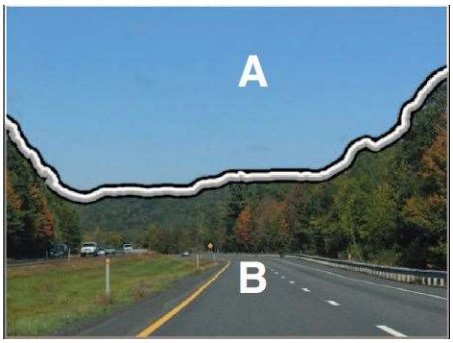


Figure 1: A sample image segmented into two regions A and B
Source: [11]

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

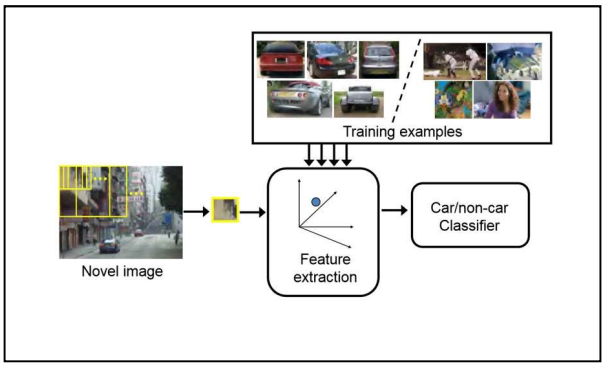


Figure 2: Main components of a sliding window detector
Source: [12]

IV. METHODOLOGY

Every object recognition algorithm must be trained to recognise an object and must know the object. For the purpose of this research work and to test our model, the chosen object of interest was a "mug". The upper section of fig 3 shows the generic model for object recognition using the sliding window approach. In this case "car" was the object of interest. The lower part of fig 3 shows our extension of the model to enable the identification of the object colour.

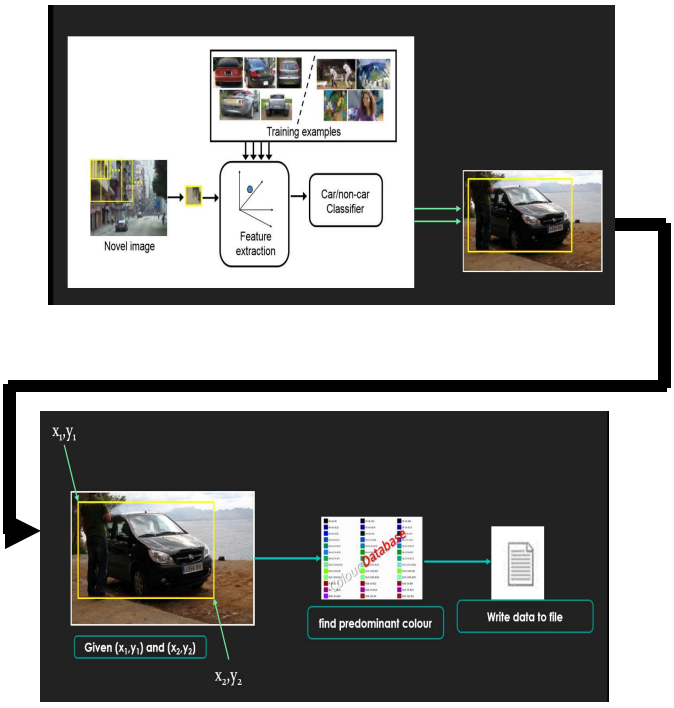


Figure 3: Model for Pixel location, Colour alignment and Automatic colour recognition

A. Dataset

The image data set of 1,950 images were obtained from several sources on the Internet and personal picture collections. The images were split into two folders, which were the "true" images and the "false" images folder. Fig 4 shows a cross section of the "true" images folder called Mugs. This folder contains pictures of mugs with different shapes, colours and sizes. Images that contained more than one mug, were cropped, such that one file had just a picture of a mug in it. These pictures were used by the algorithm to detect and select the desired mug features. It was observed that the algorithm to train a Cascade Object Detector works faster and better if the training images are in grayscale. Consequently, all the images in the "mugs" folder were converted to grey scale and stored in another folder "gray_mugs". It was also further observed that the gray_mugs images were of different dimensions/ resolutions and this posed a problem for the training algorithm. Some images were $1500\ pixels \times 1500\ pixels$ in resolution while others were 150×256. Therefore, the resolution of all the "true" greyscale images were reduced such that the resolution of each image was not more than 100×100 pixels. This reduction was done, while maintaining the aspect ratio of the images as a whole.

Figure 4: A cross section of the true image dataset

The positive training examples contained 1,160 images of mugs together with their left-to-right reflections and upside-down reflections. The negative samples had a total of 205 pictures of various dimensions and were stored in a folder called "gray_not_mugs".

B. Training

The object recognition algorithm was trained using MATLAB. The stages used for object recognition in MATLAB are designed to reject negative samples as fast as possible. The assumption is that the vast majority of windows do not contain the object of interest [14] conversely, true positives are rare, and worth taking the time to verify. To perform well, each stage in the cascade must have a low false negative rate. There are several features used in MATLAB for object recognition. The object of interest determines the feature to be chosen. It is important to choose the feature which suits the type of object detection that is needed. The "trainCascadeObjectDetector" toolbox in MATLAB, supports three types of features namely:

- Haar
- Local Binary Patterns (LBP)
- Histograms of Oriented Gradients (HOG)

The HOG features have been used for detecting objects such as people, house and cars. They are useful for capturing the overall shape of an object. The HOG features was used for this work.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The images in the "gray_mugs" folder as well as the "gray_not_mugs" folder were used to train the algorithm for object recognition. Fig 5 shows the steps involved in the training as well as how the images were converted to grey scale. There are several stages involved in training and this is either determined by the programmer or by the MATLAB based on some criteria. The default number of training stages is 20 if not specified by the programmer however the training can stop before it gets to the 20th stage, if the desired output is gotten or the module runs out of images to use for training. Increasing the number of stages may result in a more accurate detector however, this will increase the training time. The actual number of positive samples used at each stage is determined automatically by the "trainCascadeObjectDetector" function and is based on the number of stages as well as the true positive rate specified [14]. The images in gray_not_mugs folder were used to generate negative samples thus they do not contain any objects of interest, however they contain backgrounds associated with the object.

The output of the training stage in Fig 5 is an XML file called "mug_det.xml". This file contains all the features extracted during the training phase of this research work that will enable the recognition of a mug.

```
% Train for object recognition

 display(mugs);
 for i=1:length(mugs)
    display(i);
    imwrite(rgb2gray(imread(mugs(i,:))),
strcat('gmug',int2str(i),'.jpg'));
 end

for i=1:length(not_mugs)
    display('Writing not mugs');
    if(ismatrix(not_mugs(i,:)))
       imwrite(rgb2gray(imread(not_mugs(i,:))),
strcat('gnmug',int2str(i),'.jpg'));
    end
 end

mugDir = fullfile('image_data_set', 'gray_mugs', '*.jpg');
notMugDir = fullfile('image_data_set', 'gray_not_mugs');
dirOut = dir(mugDir);
img = {dirOut.name}';

% Create array of structs for the positive image samples
disp('Creating array of struct of positive samples...');
for i=1:length(img)
   image_i = imread(img{i});
   img_size = size(image_i);
   mugs_arr(i) = struct('imageFilename', img{i},
'objectBoundingBoxes', [1 1 img_size(2), img_size(1)]);
end

% Train cascade detector
disp ('training the detector...');
trainCascadeObjectDetector('mug_det.xml', mugs_arr, notMugDir);
```

Figure 4: Steps used to train for object recognition

C. Object Recognition

Fig 6 shows the function "mugDetector" written MATLAB that uses mug_det.xml as a parameter for object recognition. This was used in the recognition of a mug. The regular output of such functions is an image displayed on the screen with a yellow bounding box for each detected object of interest (see fig 7). However, one of the variables used in the function "bbox" contains some useful data. This data, which are a set of numbers ultimately represent the start pixel location of the detected object in the input image as well as the dimension of the object. The data can then be used to fetch the two pairs of pixel coordinates X_1Y_1 and X_2Y_2 (fig 7). If no object is detected in the input image, then this variable is empty. Consequently, an additional output was included in this function, to save the content of "bbox" to a text file ('object_boundary.txt') as this variable also contains the height and width of the mug detected in an input image.

```
1   function mugDetector( input_image )
2   %  mugDETECTOR Summary
3   %  This function accepts a file name as its input
4   %  and tries to detect if there is a mug in it
5
6   if(ischar(input_image))
7   img = imread(input_image);
8   end
9   [h, w] = size(img);
10  det = vision.CascadeObjectDetector('mug_det.xml');
11  bbox = step(det, img);
12  iframe = insertObjectAnnotation(img, 'rectangle', bbox, 'Mug');
13  imshow(iframe);
14  fileID = fopen('object_boundary.txt','w');
15  fprintf(fileID,'%d\t%d\t%d\t%d\n',bbox');
16  fclose(fileID);
17   end
```

Figure 5: Object recognition steps

Figure 6: Pixel location and colour identification

D. Object colour identification

In order to identify the predominant colour of a recognized object, the text file 'object_boundary.txt' which contains the location and dimensions of the recognized object was used. A model was developed to detect the predominant colour within the pair of co-ordinates supplied in the text file using the following steps:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

1. Load image A
2. Read text file 'object_boundary.txt' to get object location and dimension
3. Fetch and use X_1Y_1 and X_2Y_2 as boundary
4. Scan within X_1Y_1 and X_2Y_2 and generate it's colour histogram
5. Select the colour with the highest frequency
6. Fetch its RGB combination and thus the colour
7. Display result

V. RESULTS

The usual output of any object recognition model is an image with the identified object bounded in a yellow rectangle. This allows the user to have a visual clue of the actual location of the identified object. However, the model developed in this research work was able to identify the predominant colour of an identified object. It achieved this by identifying the predominant colour within the yellow bounding box which ultimately corresponded with the predominant colour of the object. This is because the 'object' usually occupies at least 80% of the area within the bounding box. For the purpose of this research work, "mugs"/ "cups" were identified and the various predominant colours also identified and verified.

Some of the results are presented in Fig 8 and 9. When the whole image in Fig 8(a) was passed through the model, the predominant colour using the (R, G, B) model was identified as (121, 108, 99). This code from the colour database [20] translates to the colour shown in Fig 8(b). However, with the location of the object, the predominant colour for that region was identified as (30, 98, 181) which has the presentation shown in Fig 8(c).

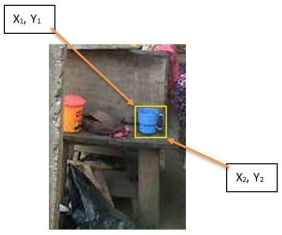


Figure 7 (a) : Test image with recognised object

(b) Predominant colour for the test image

(c) Predominant colour of the object

The result gotten from another test image is shown in Fig 9. The RGB code for the predominant colour within the bounding box was gotten as (153, 192, 77). When this colour code combination in looked up in the colour database, it was visually presented as the colour shown in in Fig 9(b). This proves to a great extent that using our model, and given any pair of co-ordinates that represent any two-dimensional shape, the predominant colour within that shape can be obtained as a set of RGB code (#, #, #), where # represents integer values ranging from zero '0' to two hundred and fifty-five '255'.

This model can therefore be used to obtain the predominant colour of any object, provided that the location of the object in the digital image can be specified as a pair of co-ordinates that stipulate the pixel location.

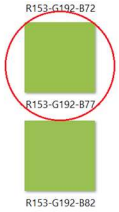


Figure 8: (a) Input image (b) Predominant colour of object

VI. SUMMARY

Visible objects in images were identified and their predominant colour recognized to minimize human error with respect to colour identification. An algorithm was developed to locate a pixel within an recognized object in an image, emphasize its colours and automatically recognize the predominant colour of the identified object in that digital image. This predominant colour was verified from a colour database. This model can be applied to several fields such as in Agriculture, (to monitor the stages in plant phenology/plant growth). In Legal systems, to correctly qualify an image.

VII. REFERENCES

[1] Y. Ramadevi, T. Sridevi, B. Poornima and B. Kalyani, "Segmentation and Object Recognition Using Edge Detection Techniques," *International Journal of Computer Science and Information Technology,* pp. 153-161, 2010.

[2] H. Whiteman, "Scientists look at 'The Dress' and answer the color conundrum," 2015. [Online]. Available: http://www.medicalnewstoday.com/articles/293967.php. [Accessed 20 May 2015].

[3] M. Ramek, "Colour vision and computer-generated images," *Journal of Physics: Conference Series,* vol. 237, no. 1, p. 012018, 2010.

[4] I. Hatem and J. Tan, *Image Analysis,* New York: Taylor and Francis, 2011, pp. 847-853.

[5] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., New Jersey: Pearson Prentice Hall, 2008.

[6] P. E. Romano, "Opthalmol," 1998.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[7] K. Choudhary, M. Pundlik and D. Choukse, "An Integrated Approach for Image Retrieval based on Content," *International Journal of Computer Science Issues,* vol. 7, no. 3, pp. 42-47, 2010.

[8] R. C. Hardie and M. M. Hayat, "Digital Image Processing," *Encyclopedia of Optical Engineering,* pp. 403-410, 13 December 2007.

[9] C. Chan, "Fundamentals of Digital Image Processing," 2015. [Online]. Available: http://www.eie.polyu.edu.hk/~enyhchan/imagef.pdf. [Accessed 6 January 2015].

[10] Y. Liu, D. Zhang, G. Lu and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition 40,* pp. 262-282, 2007.

[11] M. K. Johnson, K. Dale, S. Avidan, H. Pfister, W. T. Freeman and W. Matusik, "CG2Real: Improving the Realism of Computer Generated Images using a Large Collection of Photographs," *IEEE Transactions on Visualization and Computer Graphics,* vol. XVII, no. 9, pp. 1273 - 1285 , 2011.

[12] K. Grauman and B. Leibe, Visual Object Recognition, A Draft., 2011.

[13] M.-H. Yang, "Object Recognition," 05 September 2017. [Online]. Available: http://faculty.ucmerced.edu/mhyang.

[14] I. The MathWorks, "Object Recognition," 2014. [Online]. Available: http://www.mathworks.com/discovery/object-recognition.html. [Accessed 4 November 2014].

[15] L. Wang, J. Shi, J. Song and I.-f. Shen, "Object Detection Combining Recognition and Segmentation," in *ACCV*, 2007.

[16] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," in *International Conference on Computer Vision*, 1999.

[17] G. Csurka, C. Bray, C. Dance and L. Fan, "Visual Categorization with Bags of Keypoints," in *Workshop on Statistical Learning in Computer Vision, in conjunction with ECCV*, 2004.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, 2001.

[19] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *CVPR 2005,* 2005.

[20] N. Woods and A. B. C. Robert, "A Model for Creating Exact Colour Spectrum for Image Forensic," *University of Ibadan Journal of Science and Logics in ICT Research,* vol. 1, pp. 1-6, 2017.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Average Conjugate Gradient Method With Optimum Restart Powell For Nonlinear Function

Rana Z. Al-Kawaz
Department of Mathematics, College of Basic Education, University of Telafer, Iraq.
Email:-rana.alkawaz@yahoo.com

***Abstract*- In this article, presents a new technique to β_k parameter and check the sensitive about the scalar for Powell restart to solve unconstrained optimization functions. Global convergence with new approaches is established to minimize the selected testing functions for the conjugate gradient algorithms.. The new methods are tested on a number of benchmark functions that have been extensively used for general functions and numerical results showing the competitiveness of new methods.**

I. INTRODUCTION

We can give a generally clear definition of the unconstrained optimization problem:

$$\text{Min } f(x), \quad x \in R^n \tag{1}$$

The function can be given in form $f: R^n \to R$ their quality is general and can be found gradient $g(x)$. In the repetition k, which will be considered the previous repetition of the new step x_{k+1} of the next common gradient method:

$$x_{k+1} = x_k + \lambda_k d_k \tag{2}$$

$$d_{k+1} = \begin{cases} -g_{k+1}, & for \quad k = 0 \\ -g_{k+1} + \beta_k d_k & for \quad k \geq 1 \end{cases} \tag{3}$$

where λ_k the distance between two point; d_k is the direction of past research; define $y_k = g_{k+1} - g_k$ and β_k is a parameter for this method. Consider $\|.\|$ the Euclidean norm . The CG-algorithm use the line search, in many cases we rely on the qualifications of Wolfe Powell to find the new step [12]:

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \delta \lambda_k g_k^T d_k, \tag{4}$$

$$g_{k+1}^T d_k \geq \sigma \, g_k^T d_k, \tag{5}$$

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \delta \lambda_k g_k^T d_k, \tag{6}$$

$$\left| g_{k+1}^T d_k \right| \leq -\sigma \, g_k^T d_k \tag{7}$$

Where $0 < \delta < 0.5 \leq \sigma < 1$. Equations [(4) and (5)] - [(6) and (7)] are called "Weak Wolfe Powell"-"Strong Wolfe Powell" qualifications [12], respectively. CG algorithms vary according to the formula of the correlation coefficient β_k used in the formula and on the basis of which these algorithms are classified [1]. We include some key formulas that can be the basis for the rest of the derived formulas:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE I

COMPARATIVE BETWEEN THE CLASSICAL FORMULA β_k

Advantage / Strong Convergence Properties		Advantage / Better Computational Performances	
$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$	Fletcher-Reeves (FR) [13]	$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k}$	Polak-Ribiere-Polyak (PRP) [4]
$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T d_k}$	Dai-Yuan (DY) [16]	$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}$	Hestenes and Stiefel (HS) [7]
$\beta_k^{CD} = -\frac{g_{k+1}^T g_{k+1}}{g_k^T d_k}$	Descent Fletcher (CD) [14]	$\beta_k^{LS} = -\frac{g_{k+1}^T y_k}{g_k^T d_k}$	Liu and Storey (LS) [17]

. Recently, many researchers have developed these classical formulas in order to obtain results and optimal convergence in the solution; (Rivaie_Mustafa_Ismail_Leong) (RMIL) method [9], (Syazni_Rivaie_Mustafa_Ismail) (SRMI) method [15], (Norrlaili_Rivaie_Mustafa_Ismail) (NRMI) method [11] and (Rivaie_Abashar_Mustafa_Ismail) (RAMI) method [8].

$$\beta_k^{RMIL} = \frac{g_{k+1}^T y_k}{\|d_k\|^2} \tag{8}$$

$$\beta_k^{NRMI} = \frac{g_{k+1}^T y_k}{g_k^T (g_{k+1} - d_k)} \tag{9}$$

$$\beta_k^{SRMI} = \frac{\frac{g_{k+1}^T y_k}{\|g_k\|^2} + \frac{g_{k+1}^T y_k}{g_k^T (g_{k+1} - d_k)}}{2} \tag{10}$$

$$\beta_k^{RAMI} = \frac{g_{k+1}^T \left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k \right)}{d_k^T (d_k - g_{k+1})} \tag{11}$$

The article is divided into major parts as follows:
1. General introduction of the conjugation methods showing their composition.
2. The coefficients of CG coefficients as suggested in this article with the writing of its algorithm.
3. The theories of convergence analysis are preceded by the theory of sufficient descent of the new research direction.
4. Calculate the results of the new algorithm and include them in tables to highlight their efficiency.
5. Finally discussing the results of the new algorithm with the results of the Powell scale of the changer restart.

II. AVERAGE OF THE NEW β_k CG-METHOD

A. Clarify the new idea

During the reading of many of the research and practical experiences I noticed an idea put forward to the researchers (Syazni-Rivaie-Mustafa-Ismail) (SRMI) and its experience on the practical side gave me good results when circulating three parameters of the parameters presented in this article can also be circulated for more to take advantage of the features of each parameter and integrate these features to give a strong parameter in both

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

theoretical and practical aspects for the associated gradient algorithms that solve unconstrained optimization problems.

Let's take a look at the new idea proposed in this section:

$$\beta_k^{New} = \frac{\frac{g_{k+1}^T y_k}{\|g_k\|^2} + \frac{g_{k+1}^T y_k}{(u\ g_k^T g_{k+1} - v\ g_k^T d_k)} - \frac{g_{k+1}^T y_k}{g_k^T d_k}}{3} \quad (12)$$

$$\beta_k^{New} = \text{Average of } (\beta_k^{PR}, \beta_k^{NRMI}, \beta_k^{LS})$$

$$d_{k+1} = -g_{k+1} + \beta_k^{New} d_k \quad (13)$$

First new algorithm is a combination of the preferred characteristics of the parameters ($\beta_k^{PR}, \beta_k^{NRMI}, \beta_k^{LS}$) in a statistical manner by taking the resulting rate of their mathematical properties. The new algorithm offered its best when adding the Powell feature to the updated retrieval by replacing the amount of the value of the condition in it. The second algorithm for this article and the two algorithms converged to the optimal solution on a regular basis.

Powell [1] noted that the classical algorithms presented in the first chapter may fail to achieve a decrease of function and derivative and therefore may not give us a downward search trend towards reaching the point of diminution of the function (optimal point) and because of the performance of the parameter β_k .So the condition that corresponds to the process of re-trajectory get off again where condition [6]:

$$\left|g_{k+1}^T g_k\right| > 0.2 \left\|g_{k+1}\right\|^2 \quad (14)$$

and by developing this requirement to

$$\left|g_{k+1}^T g_k\right| > 0.9 \left\|g_{k+1}\right\|^2 \quad (15)$$

Gave outstanding results in computational performance and the second algorithm was with the new parameter β_k^{New}.

B. Outline of The New 1 CG-Method

Step 1: The input as values $x_1 \in R^n$; ($\varepsilon > 0$); (k) is the algorithm's index.

Step 2: First iteration will be using k=1; set $d_k = -g_k$.

Step 3: Calculate the distance of the next step using the strong line search terms Wolfe Powell (6) - (7).

Step 4: Calculate the new point as $x_{k+1} = x_k + \lambda_k d_k$ and compute f , g ,use $y_k = g_{k+1} - g_k$

Step 5: If the Powell restart (15) satisfied, put it: $d_{k+1} = -g_{k+1}$, else set

$d_{k+1} = -g_{k+1} + \beta_k^{New} d_k$ (β_k^{New} in (12)) and go to Step 2.

Step 6: If $\left\|g_{k+1}\right\| < \varepsilon$, stop else set k=k+1 go to Step 3.

C. Outline of The New 2 CG-Method

Step 1: The input as values $x_1 \in R^n$; ($\varepsilon > 0$); (k) is the algorithm's index.

Step 2: First iteration will be using k=1; set $d_k = -g_k$.

Step 3: Calculate the distance of the next step using the strong line search terms Wolfe Powell (6) - (7).

Step 4: Calculate the new point as $x_{k+1} = x_k + \lambda_k d_k$ and compute f , g ,use $y_k = g_{k+1} - g_k$

Step 5: If the Powell restart (15) satisfied, put it: $d_{k+1} = -g_{k+1}$, else set

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$d_{k+1} = -g_{k+1} + \beta_k^{New} d_k$ (β_k^{New} in (12)) and go to Step 2.

Step 6: If $\|g_{k+1}\| < \varepsilon$, stop else set k=k+1 go to Step 3.

III. Theoretical Properties for the New CG-Methods

In this section, we focus on rapprochement behavior of the β_k^{New} using inexact line search. Now, we write appropriate hypotheses to prove the convergence of the two new comprehensive algorithms to arrive at the optimal point plus the evidence of sufficient regression.

A. *Assumption.*[2]

If it were f restricted from bottom to level set $L_{x_0} = \{x \in R^n : f(x) \le f(x_0)\}$; in some neighborhood N of this set and $f(x)$ is continuously differentiable and its gradient $g(x)$ is Lipschitz continuous in the same set , namely, there exists L> 0 such that:

$$\|g(x) - g(y)\| \le L\|x - y\| \qquad \text{for all x, y} \in L_{x_0} \tag{16}$$

B. *Assumption.*[2] The level set L_{x_0} is compact.

C. *Theorem*

Suppose that Assumption (A) , (B) holds. If there is a constant $\gamma > 0$ such that $\gamma \le \|g_k\| \le \bar{\gamma}$, for all $k \ge 0$. If λ_k they are obtained through Wolfe Powell's strong conditions and d_{k+1} is given in (13) of the new β_k^{New} CG-method (12) and the scalar (0<u<v<1), then the new method has sufficient descent directions i.e.,

$$g_{k+1}^T d_{k+1} \le -c\,\|g_{k+1}\|^2; \quad c > 0 \tag{17}$$

Proof:

Multiplying the new search direction of (13) by g_{k+1} and by substituting β_k^{New} of (12) yields:

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \beta_k^{New} g_{k+1}^T d_k$$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \left(\frac{\frac{g_{k+1}^T y_k}{\|g_k\|^2} + \frac{g_{k+1}^T y_k}{(u\ g_k^T g_{k+1} - v\ g_k^T d_k)} - \frac{g_{k+1}^T y_k}{g_k^T d_k}}{3} \right) g_{k+1}^T d_k$$

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{3\|g_k\|^2} g_{k+1}^T d_k + \frac{g_{k+1}^T y_k}{3(u\ g_k^T g_{k+1} - v\ g_k^T d_k)} g_{k+1}^T d_k - \frac{g_{k+1}^T y_k}{3 g_k^T d_k} g_{k+1}^T d_k$$

Through the Wolf search line we can get the next inequality ($g_{k+1}^T d_k \le -\sigma\, g_k^T d_k$) then get:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{3\|g_k\|^2}(-\sigma\, g_k^T d_k) + \frac{g_{k+1}^T y_k}{3(u\ g_k^T g_{k+1} - v\ g_k^T d_k)}(-\sigma\, g_k^T d_k) - \frac{g_{k+1}^T y_k}{3 g_k^T d_k}(-\sigma\, g_k^T d_k)$$

And through algebraic operations:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}\|^2 + \frac{2}{3}\sigma\, g_{k+1}^T y_k + \frac{g_{k+1}^T y_k}{3(u\ \omega\|g_{k+1}\|^2 + v\ \|g_k\|^2)} \sigma \|g_k\|^2$$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

To sum up the amount $g_{k+1}^T y_k$ using Powell's condition for restart when $\omega = 0.9 \; or \; 0.2$

$$g_{k+1}^T y_k = g_{k+1}^T g_{k+1} - g_{k+1}^T g_k < (1-\omega)\left\|g_{k+1}\right\|^2$$

$$g_{k+1}^T d_{k+1} \le -\left\|g_{k+1}\right\|^2 + \frac{2}{3}\sigma(1-\omega)\left\|g_{k+1}\right\|^2 + \frac{(1-\omega)\left\|g_{k+1}\right\|^2}{3(u\ \omega\left\|g_{k+1}\right\|^2 + v\ \left\|g_k\right\|^2)}\sigma\left\|g_k\right\|^2$$

$$g_{k+1}^T d_{k+1} \le -\left(1 + \frac{2}{3}\sigma(1-\omega) + \frac{(1-\omega)}{3(u\ \omega\left\|g_{k+1}\right\|^2 + v\ \left\|g_k\right\|^2)}\sigma\left\|g_k\right\|^2\right)\left\|g_{k+1}\right\|^2$$

Where $c = 1 + \frac{2}{3}\sigma(1-\omega) + \frac{(1-\omega)}{3(u\ \omega\left\|g_{k+1}\right\|^2 + v\ \left\|g_k\right\|^2)}\sigma\left\|g_k\right\|^2 > 0$

$$g_{k+1}^T d_{k+1} \le -c\left\|g_{k+1}\right\|^2$$

Now the two new algorithms have achieved the condition of sufficient descent. We can move on to the other important condition, the condition of convergence, so that our algorithm complements its theoretical conditions in the correct manner, as in the following theory:

D. Lemma. (Zoutendijk Condition).[5]

Let's give that option of Assumption (A), (B) satisfied. Suppose that any CG style is in the form $x_{k+1} = x_k + \lambda_k d_k$ where d_{k+1} is a descent direction and λ_k satisfies the Strong Wolfe-Powell line search conditions in (6)-(7) . Then we have that:

$$\sum_{k\ge 0}\frac{(g_k^T d_k)^2}{\left\|d_k\right\|^2} < +\infty \qquad (18)$$

E. Theorem.

Suppose Assumption (A), (B) is correct. Consider the new CG method specified in the (12), (13) with β_k^{New}, if λ_k is obtained by an inexact line search (6)-(7) which satisfied the sufficient descent condition (18). then:

$$\liminf_{k\to\infty}\left\|g_k\right\| = 0 \qquad (19)$$

Proof:

We now establish the theory by contradiction and assume some constants $\gamma > 0$ such that $\left\|g_k\right\| \ge \gamma$ for all $k \ge 0$. The compactness of the level set L_{x_0} implies that there exists a constant $\bar{\gamma} > 0$ such that $\left\|g_k\right\| \le \bar{\gamma}$. either $g_k = 0$ for some k or $\liminf_{k\to\infty}\left\|g_k\right\| = 0$

Because the case of descent condition holds, we have $\left\|d_k\right\| \ne 0$ .using the Lipchitz condition

$$\left\|y_k\right\| = \left\|g_{k+1} - g_k\right\| \le L\left\|s_k\right\|$$

then,

$$\left|\beta_k^{New}\right| = \left|\frac{\frac{g_{k+1}^T y_k}{\left\|g_k\right\|^2} + \frac{g_{k+1}^T y_k}{(u\ g_k^T g_{k+1} - v\ g_k^T d_k)} - \frac{g_{k+1}^T y_k}{g_k^T d_k}}{3}\right|$$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$\left|\beta_k^{New}\right| \le \frac{1}{3}\left\|g_{k+1}\right\|\left\|y_k\right\|\left(\frac{1}{\left\|g_k\right\|^2}+\frac{1}{(u\left\|g_k\right\|\left\|g_{k+1}\right\|+v\left\|g_k\right\|\left\|d_k\right\|)}+\frac{1}{\left\|g_k\right\|\left\|d_k\right\|}\right)$$

$$\left|\beta_k^{New}\right| \le \frac{1}{3}\Gamma L\left\|s_k\right\|\left(\frac{1}{\gamma^2}+\frac{\alpha}{(u\,\alpha\,\gamma\,\Gamma+v\ \gamma\left\|s_k\right\|)}+\frac{1}{\gamma\left\|s_k\right\|}\right)$$

D as knew in assumption (A) and

$$\left|\beta_k^{New}\right| \le \frac{1}{3}\Gamma LD\left(\frac{1}{\gamma^2}+\frac{\alpha}{(u\,\alpha\,\gamma\,\Gamma+v\ \gamma D)}+\frac{1}{\gamma D}\right)\equiv E$$

So the new direction ,

$$\left\|d_{k+1}^{New}\right\| \le \left\|g_{k+1}\right\|+\left|\beta_k^{New}\right|\left\|s_k\right\|$$

$$\le \Gamma + ED$$

This implies

$$0 < \sum_{k=1}^{\infty}\frac{(g_k^T d_k)^2}{\left\|d_k\right\|^2} < \infty$$

$$\sum_{k=1}^{\infty}\frac{\left\|g_k\right\|^4}{\left\|d_k\right\|^2} < \frac{1}{c^2}\frac{(g_k^T d_k)^2}{\left\|d_k\right\|^2} < \infty$$

The new algorithm has achieved global convergence. We move on to the practical side of these proposed algorithms as in the following section:

IV. Numerical Results

This section includes the performance of new methods on a set of test problems. The codes are written in Fortran and in the double precision calculation. Using a computer, all tests are performed. Our experiments are conducted on a set of 35 nonlinear nonlinear cases that can be derived from a transducer. These problems contribute to the CUTE test and are detailed in Dolan, Moré [3] and Andrei [10]. All algorithms apply the same standard stop, s.t. $\left\|g_k\right\|_\infty \le 10^{-6}$. Algorithms comparisons are presented in the following context. The digit codes are written in Fortran and integrated with Visual Fortran with the physical components of the Intel Pentium 4 and 1.86 GHz. Numerical experiments with number of variables n = 50, 400,750.

TABLE II

NUMERICAL RESULTS FOR NEW 1 ALGORITHM AGAINST (PR & RMIL) WHEN USE THE (50<N<750) DIMENSIONS AND WITH $(\omega = 0.2)$

Function	New 1 CG-Algorithm / $\omega = 0.2$ NOI/NOFG/TIME	Best against	PR (Classical CG-lgorithm) NOI/NOFG/TIME	RMIL NOI/NOFG/TIME
Freudenstein & Roth - FREUROTH	1372/38085/0.46	PR/RMIL	2459/14713/0.28	6003/6129/0.37
Trigonometric	68/124/0.01		63/123/0.03	68/127/0.05
Extended Rosenbrock SROSENBR	287/617/0.02	PR/RMIL	291/605/0.03	6003/6012/0.35
Extended White & Holst	61/116/0.00	PR/RMIL	90/151/0.00	3690/7379/0.27
Extended Beale	56/105/0.00	PR/RMIL	92/155/0.08	6003/6009/0.32
Penalty	13/33/0.00	RMIL	12/32/0.00	29/66/0.00

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Perturbed Quadratic	688/1122/0.01	RMIL	576/926/0.04	4705/7839/0.37
Raydan 1	789/1321/0.18	RMIL	594/963/0.16	4070/5741/0.93
Raydan 2	12/27/0.01		12/27/0.00	12/27/0.00
Diagonal 1	348/738/0.00	PR/RMIL	2209/2551/0.06	6003/6019/0.35
Diagonal 2	693/1289/0.25	PR	4086/4191/0.88	532/882/0.19
Diagonal 3	2882/73260/20.32	PR /RMIL	3767/97736/24.53	6003/6378/1.30
Hager	534/14336/3.65	RMIL	324/6638/1.55	4027/4299/1.12
Generalized Tridiagonal 1	74/142/0.02	RMIL	73/140/0.00	6003/6181/0.35
Extended Tridiagonal 1	66/128/0.00	PR/RMIL	118/183/0.00	46/97/0.01
Extended Three Expo Terms	42/69/0.00	RMIL	34/60/0.00	96/179/0.05
Generalized Tridiagonal 2	160/250/0.00	RMIL	136/234/0.00	4178/4512/0.37
Diagonal 4	33/63/0.00		12/24/0.00	12/24/0.00
Diagonal 5	6003/11202/1.12		2469/3430/0.49	6003/8026/0.91
Extended Himmelblau	42/77/0.00	RMIL	33/69/0.02	149/301/0.00
Generalized PSC1	1428/2738/0.05	PR/RMIL	1771/2722/0.09	6003/8073/0.34
Extended PSC1	26/55/0.00	PR/RMIL	34/71/0.00	34/66/0.00
Extended Powell	33/83/0.00	PR /RMIL	46/94/0.00	4064/4142/0.56
Extended BD1	54/110/0.00	RMIL	54/115/0.00	110/226/0.00
Extended Maratos	56/91/0.21	PR/RMIL	58/97/0.54	6003/6028/28.98
Extended Cliff	116/290/0.02	PR/RMIL	4018/4060/0.36	6003/8016/0.94
Quadratic Diagonal Perturbed	682/1219/0.03	RMIL	390/698/0.05	6003/8120/0.40
Extended Wood WOODS (CUTE)	4363/11736/1.14	PR/RMIL	5104/16799/3.74	6003/8307/1.05
Extended Hiebert	41/83/0.00	RMIL	40/84/0.00	68/137/0.03
Quadratic QF1	6003/10579/0.19		6003/6075/0.38	6003/6028/0.39
Extended Quadratic Penalty QP1	3993/6304/0.16	PR/RMIL	4415/4585/0.33	6003/6030/0.35
Extended Quadratic Penalty QP2	229/450/0.00	RMIL	160/287/ 0.05	6003/6024/0.39
Quadratic QF2	2615/4244/0.10	PR /RMIL	3013/3913/0.26	6003/6206/0.33
Extended EP1	48/87/0.02	RMIL	38/74/0.00	6003/6125/0.37
Extended Tridiagonal 2	50/97/0.00	RMIL	44/86/0.00	6003/6128/0.32

TABLE III

NUMERICAL RESULTS FOR NEW 1 ALGORITHM AGAINST (PR & RMIL) WHEN USE THE (50<N<750) DIMENSIONS AND WITH $(\omega = 0.2)$

Function	New 2 CG-Algorithm / $\omega = 0.9$ NOI/NOFG/TIME	Best against	PR (Classical CG-Algorithm) NOI/NOFG/TIME	RMIL NOI/NOFG/TIME
Freudenstein & Roth -	274/6187/0.22	PR/RMIL	2459/14713/0.28	6003/6129/0.37

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

FREUROTH				
Trigonometric	61/121/0.03	PR /RMIL	63/123/0.03	68/127/0.05
Extended Rosenbrock SROSENBR	212/455/0.02	PR /RMIL	291/605/0.03	6003/6012/0.35
Extended White & Holst	64/120/0.00	PR/RMIL	90/151/0.00	3690/7379/0.27
Extended Beale	55/103/0.03	PR/RMIL	92/155/0.08	6003/6009/0.32
Penalty	13/33/0.00	RMIL	12/32/0.00	29/66/0.00
Perturbed Quadratic	586/823/0.05	RMIL	576/926/0.04	4705/7839/0.37
Raydan 1	574/840/0.12	PR /RMIL	594/963/0.16	4070/5741/0.93
Raydan 2	12/27/0.00		12/27/0.00	12/27/0.00
Diagonal 1	308/599/0.01	PR/RMIL	2209/2551/0.06	6003/6019/0.35
Diagonal 2	436/1274/0.30	PR/RMIL	4086/4191/0.88	532/882/0.19
Diagonal 3	3007/6352/24.39	PR /RMIL	3767/97736/24.53	6003/6378/1.30
Hager	407/9867/2.53	RMIL	324/6638/1.55	4027/4299/1.12
Generalized Tridiagonal 1	67/141/0.00	PR/RMIL	73/140/0.00	6003/6181/0.35
Extended Tridiagonal 1	43/88/0.00	PR/RMIL	118/183/0.00	46/97/0.01
Extended Three Expo Terms	38/64/0.02	RMIL	34/60/0.00	96/179/0.05
Generalized Tridiagonal 2	122/197/0.00	PR /RMIL	136/234/0.00	4178/4512/0.37
Diagonal 4	36/69/0.00		12/24/0.00	12/24/0.00
Diagonal 5	4504/8761/1.41	RMIL	2469/3430/0.49	6003/8026/0.91
Extended Himmelblau	30/55/0.00	PR /RMIL	33/69/0.02	149/301/0.00
Generalized PSC1	1093/2114/0.11	PR/RMIL	1771/2722/0.09	6003/8073/0.34
Extended PSC1	20/44/0.00		34/71/0.00	34/66/0.00
Extended Powell	34/84/0.00	PR /RMIL	46/94/0.00	4064/4142/0.56
Extended BD1	52/122/ 0.01	PR /RMIL	54/115/0.00	110/226/0.00
Extended Maratos	46/84/0.41	PR /RMIL	58/97/0.54	6003/6028/28.98
Extended Cliff	91/200/0.01	PR/RMIL	4018/4060/0.36	6003/8016/0.94
Quadratic Diagonal Perturbed	343/681/0.03	PR/RMIL	390/698/0.05	6003/8120/0.40
Extended Wood WOODS (CUTE(	3978/40129/8.34	PR/RMIL	5104/16799/3.74	6003/8307/1.05
Extended Hiebert	37/78/0.02	PR/RMIL	40/84/0.00	68/137/0.03
Quadratic QF1	5036/9248/0.50	PR/RMIL	6003/6075/0.38	6003/6028/0.39
Extended Quadratic Penalty QP1	3839/5240/0.35	PR /RMIL	4415/4585/0.33	6003/6030/0.35
Extended Quadratic Penalty QP2	245/465/0.02	RMIL	160/287/0.05	6003/6024/0.39
Quadratic QF2	4320/6145/0.36	RMIL	3013/3913/0.26	6003/6206/0.33
Extended EP1	38/120/0.00	RMIL	38/74/0.00	6003/6125/0.37
Extended Tridiagonal 2	46/86/0.00	RMIL	44/86/0.00	6003/6128/0.32

Through Table (we note that the new algorithm is first
1- Outperforms the PR algorithm 16 times for (NOI & NOFG & TIME) out of 35 test functions.
2- Outperforms the RMIL algorithm 29 times for (NOI & NOFG & TIME) out of 35 test function.

In Table (2) we note that the new algorithm can perform better than the first by comparing the following:
1 - Outperforms the algorithm of 23 times for (NOI & NOFG & TIME) of the 35 test function.
2 - Outperforms the RMIL algorithm 32 times for (NOI & NOFG & TIME) out of 35 test function.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. Conclusions

The new CG algorithms are applied to a number of standard functions that are average (PR & LS & RMIL) and compared with their classic classics. Therefore, we conclude from the tables given in the previous chapter. The efficiency of the second algorithm with the update of Powell's restart condition is stronger and more efficient Of the first algorithm based on the old condition and can continue in the future by making more algorithms based on the creation of the rate of a number of old classical algorithms to strengthen the efficiency of the unit and the other and increase the advantages of its performance in the solution of unconstrained optimization problems.

References

[1] A. Y. Al-Bayati and R. Z. AL-Kawaz, "A New Hybrid Global Convergent WC-FR Conjugate Gradient-Method with Relaxed Conjugacy Condition", Association for the Advancement of Modelling and Simulation Techniques in Enterprises, vol. 2, no. 1, 2013.
[2] C. Gilbert and J. Nocedal, "Global Convergence Properties of Conjugate Gradient Methods for Optimization", SIAM Journal on Optimization, vol. 2, pp. 21–42, 1992.
[3] E. D. Dolan and J. J. Moré, "Benchmarking Optimization Software with Performance Profiles", Math. Program, vol. 91, pp.201–213, 2002.
[4] E. Polak and G. Ribière, "Note Sur la Convergence de Directions Conjuguée", Rev. Francaise Informat Recherche Operationelle, 3e Année, vol. 16, pp. 35–43, 1969.
[5] G. Zoutendijk, "Nonlinear Programming, Computational Methods in Integer and Nonlinear Programming", North-Holland Amsterdam, pp. 37–86, 1970.
[6] M. J. D. Powell, "Restart procedures for the conjugate gradient method", Mathematical Program., vol. 12, pp. 241–254 ,1977.
[7] M. R. Hestenes and E.L. Stiefel, "Methods of conjugate gradients for solving linear systems" J. Res. Natl. Bur. Stand., vol. 49, pp. 409–436,1952.
[8] M. Rivaie, A. Abashar, M. Mamat and I. Mohd, "The Conjugence Properties of A New Type of Conjugate Gradient Methods", App. Math. Science 8, no. 1, pp. 33-44, 2014.
[9] M. Rivaie, M. Mamat, W. J. Leong and M. Ismail, "A New Conjugate Gradient Coefficient For Large Scale Nonlinear Unconstrained Optimization", Int. Journal of Math. Analysis, vol. 6, no. 23, pp. 1131-1146, 2012,.
[10] N. Andrei, "An Unconstrained Optimization Test Functions Collection", Advanced Modeling and Optimization, vol. 10, no. 1, pp.147-161, 2008.
[11] N. Shapiee, M. Rivaie and M. Mamat, "A New Classical Conjugate Gradient Coefficient With Exact Line Search", AIP Conf. Proc., pp. 1739,020082, 2016, doi: 10.1063/1.4952562.
[12] P. Wolfe, "Convergence Conditions for Ascent Methods", SIAM Rev., vol. 11, pp. 226–235, 1969.
[13] R. Fletcher and C. Reeves, "Function Minimization by Conjugate Gradients", Computer Journal, vol. 7, pp. 149–154, 1964.
[14] R. Fletcher, "Practical Methods of Optimization: Unconstrained Optimization", Wiley, New York , vol. 1,1987.
[15] S. Shoid, M.Rivaie and M. Mamat, "A Modification of Classical Conjugate Gradient Method Using Strong Wolfe Line Search" AIP Conf. Proc. pp. 1739, 020071, 2016 , doi: 10.1063/1.4952562.
[16] Y. H. Dai and Y. Yuan, "A Nonlinear Conjugate Gradient Method With a Strong Global Convergence Property", SIAM Journal of Optimization, vol. 10, pp. 177–182,1999.
[17] Y. Liu and C. Storey, "Efficient Generalized Conjugate Gradient Algorithms", part 1: theory. JOTA, vol. 69, pp. 129–137, 1991.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Establishing Secured Enterprise Network Routing protocols by using DMVPN.

K. Sandhya,
SNIST,
Yamnampet, Ghatkesar, Hyderabad
Telangana, Hyderabad-501301.
sandy.3527@gmail.com

V.Kakulapati
SNIST,
Yamnampet, Ghatkesar, Hyderabad
Telangana, Hyderabad-501301.
vldms@yahoo.com

Abstract---*A DMVPN (Dynamic Multipoint Virtual Private Network)is a network with meshed VPN connectivity. It eliminates limitations observed in the traditional VPN connections. It enhances the network deployment by having quick implementation of secure connections between remote sites \ branch offices (Spoke-to-Spoke) dynamically without routing all the traffic through the main branch (Hub).DMVPN is a technology that integrates different concepts such as mGRE (multipoint Generic Routing Encapsulation), NHRP (Next-Hop Resolution Protocol), Routing and IPsec. These together provide solution that allows the end users to communicate effectively through static hub-to-spoke and the dynamically created spoke-to-spoke IPSec tunnels. In this Research paper, we discussed the DMVPN developments and usage of dynamic routing protocols (OSPF, EIGRP), also simulated the static Hub-Spoke tunnel, dynamic Spoke–Spoke tunnels to ascertain the best suitable routing protocol while using in corporate / enterprise network. The simulation is carried out in a Graphic Network Simulator (GNS3).*

Keywords: DMVPN, mGRE, NHRP, NHS, IPSec, GNS3, EIGRP, OSPF.

1. INTRODUCTION

Rapid development in communication industry across globe has developed a need to secure every communication with their peers in all corporate and defence applications. Virtual Private Networks which have been used in industry are serving the purpose of secure communication using improved encryption technologies but are complex in configuring while number of branches is increasing for the organization. To eliminate the complex form of configuration and reduce the cost of O&M (Operation and Maintenance), DMVPN solution is utilized. It supports Static Hub-to-Spoke connectivity, Dynamic Spoke to spoke connectivity. It does need no additional configuration to route traffic directly between spokes and eliminates hub in the data path. The DMVPN has been developed in three phases.

a. In phase1 Hub-to-Spoke model, this is similar to point-to-point VPN, static tunnels will be present between the Hub and Spoke, but the hub learns about spokes dynamically.

b. In phase2, it is a dynamic Spoke-to-Spoke communication, channels are created on requirement.

c. In phase3 is a Spoke -to-Spoke model with enhanced features i.e., Hub and Spoke ratio.

DMVPN is built with one static address configured Hub and remaining can be dynamically addressed spoke sites. Where as in traditional VPN, the address of Hub and all Spokes must be static, and each spoke addition would need additional configuration in every spoke of the network. It is cumbersome job to the network engineer to implement, operate and maintain the traditional VPN. In a network of 25 sites, only 24 Tunnels are required in DMVPN instead of 600 to have fully meshed VPN network. It facilitates the dynamic creation of VPN tunnels while Interesting traffic is initiated for the respective remote site by learning its address from Hub.

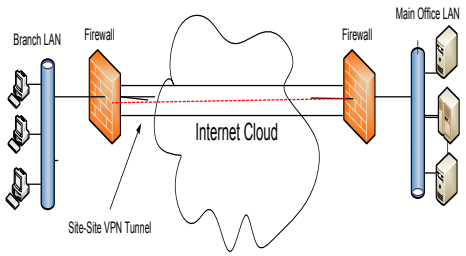


Fig 1: Secured Point to Point VPN Tunnel

DMVPN is scalable IPsec VPN that allows expanding the network to very large IPsec VPN Network. It needs no extra configuration on existing equipment but only need to add the extra piece of device on new site. It has permanent VPN tunnels to Hub site from every remote site and temporary VPN tunnels are built and vanished once data transaction is finished. Hub Router would have key role in having dynamic creation of tunnels between multiple spokes. Branch office routers built permanent tunnels to Hub Router.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

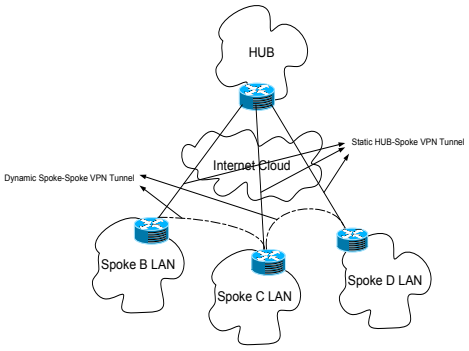


Fig 2: Dynamic Multipoint VPN Hub-Spoke model

1.1. DMVPN Phases in detail:

1.2.

Table 1 : Development Phases of Dynamic Multipoint Virtual Private Network

I Phase	II Phase	III Phase
1. IOS Version 12.2(13)T 2. Hub and Spoke performance. 3. mGRE or p-pGRE interfaces on spokes, mGRE on hubs. 4. Hubs would have interpreted and less significantformation. 5. Customer Premises Equipment (CPEs) can be enabled with Dynamic Address. 6. Routing Protocols and Multicast is supported. 7. Summarization can happen on Hubs and Spokes need not have full routing table. 8. Deployment can happen with no touch.	1. IOS Version 12.3(4)T. IWAN 1.0 2. Spoke to Spoke performance. 3. mGRE interfaces on spokes as well. 4. Data traffic through hubs is reduced by having direct spoke to spoke tunnel. 5. Daisy Chain is used in interconnection of hubs. 6. No summarization is done but full routing table at Spokes. 7. Spoke triggers the spoke to spoke tunnel. 8. Limitations of routing protocol.	1. IOS Version 12.4(6)T. IWAN 2.0 2. Greater Scaling and options for more network designs. 3. Spoke to Hub ratio would be the same. 4. Daisy Chain is not required to use in interconnecting the hubs. 5. It can summarize, and spokes does not require full routing table. 6. Hubs can trigger the spoke to spoke tunnel. 7. Limitations of Routing protocol are reduced. 8. NHRP routes andsubsequently hops available in RIB 15.2(1)T and later IOS versions.

1.3. DMVPN features:

- Dynamic VPN tunnels are triggered between specific remote sites while interesting traffic is initiated for a specific remote site host \ IP subnet.
- Existing device configurations are not needed to be modified while new remote sites are established in case of organization expansions or mergers between organizations.
- Possible to have partial \ fully meshed connectivity with simple hub and spoke configurations.
- It also supports in having less number of Static Internet (Public) IPs by having IPsec tunnels built with dynamically addressed spokes.
- MPLS and VRFs can be used along with this DMVPN with no complexity.
- IPSec encryption can be used as optional i.e., it works without encryption as well.
- IoT devices can also be used and it works well with this lower end spoke devices too.

2. RELATED WORK:

N. Angelescu.et al[8] simulated and proposed that , how DMVPN is alternative to the traditional VPN solutions in terms of efficiency , and delay simulation is carried in the GNS3 network simulation software .

Ruttajan kuniene. et al[10] analyzed the different routing protocols influence on Quality of Service (QoS) in DMVPN implementation.

Bilal Ashfaq Ahmed.et al[11] In this paper the author discussed more about how GNS3 is complementary tool that allows simulation of complex networks like real time lab implementation . Their aim is show that GNS3 is an educational tool that is used to test with different technologies without using real equipments before implementing on real equipments. To know how much complex networks can be simulated in GNS3 author compared the two different vpn techniques MPLS VPN and IPsec VPN are discussed.

3. DMVPN TECHNOLOGIES

DMVPN is built with different suite of protocols for high security of the data passed across the network. The components include the below:

- mGRE (Multipoint GRE)
- (NHRP) Next-Hop Resolution Protocol
- IPsec encryption – (Optional)
- RPs (Routing Protocols)

3.1. Multipoint GRE (mGRE):

GRE (Generic Routing Encapsulation)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

protocol support in having point to point encapsulation. While this GRE supports in having static Hub and Spoke tunnel connectivity. It would be difficult to have so many interfaces for Spoke to Spoke tunnels between every site while organisation sites\branches are increasing. Thus, m-GRE helps in reducing the number of tunnel interfaces by having terminated on single GRE interface.

- GRE\IPsec Tunnels and endpoints in multiple are supported by single GRE interface.
- Dynamic tunnel creations are supported through mGRE.
- Complexity of configuration and size are drastically reduced.
- Comparison of GRE over IPsec and DMVPN is given below in table.
- mGRE supports Hub and Spoke connectivity over single GRE interface on Hub.
- mGRE does also support in spoke to spoke connectivity to have multiple tunnels over single GRE interface.

Commands used in Simulation on Virtual Interface Tunnel0:

```
tunnel source 15.0.0.1
tunnel mode gre multipoint
tunnel key 0
tunnel protection ipsec profile dmvpnprof
```

Table 2: Features of GRE over IPSec and DMVPN

Features	3rd party Compatibility	Dynamically Addressed Spoke	Dynamic Routing	Dynamic Spoke to Spoke tunnel	QoS
GRE over IPSec	Yes	-	Yes	-	Yes
DMVPN	-	Yes	Yes	Yes	Yes

Features	Public Transport	IPv6	IP Multicast	NAT	VRF
GRE over IPSec	Yes	Yes	Yes	Yes	Yes
DMVPN	Yes	Yes	Yes	Yes	Yes

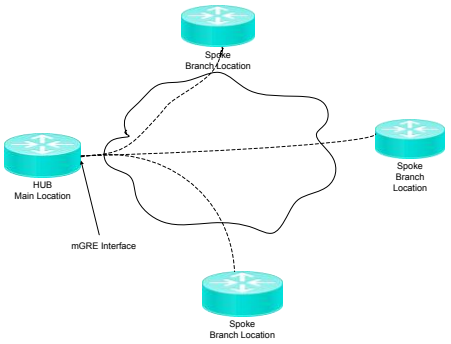


Fig 3. Hub and Spoke mGRE Tunnel Topology

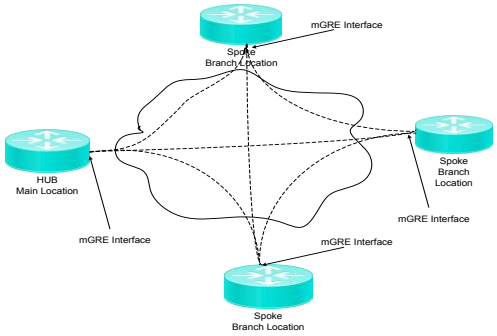


Fig 4. Spoke to Spoke mGRE Tunnel Topology

3.2. NHRP:

In DMVPN, it is needed to run Next Hop Resolution Protocol (NHRP) to have a client-server model, where Hub Router acts as Server and remaining Spoke routers act as Clients. This is like oath protocol so to allocate a SHC (subsequently Hop Client) to energetically register with SHSs (subsequently Hop Servers). Through the DMVPN device the SHC is the spoke router and the SHS is the hub router. Previously all spokes are scheduled with hub; the spoke routers can determine other spoke routers dynamically contained by the similar NBMA network.

- It emulates NBMA (Non-Broadcast Multi Access) networks on ATM and its standards.
- It creates aligning database of VPN tunnel interface to Internet (open) Interface addresses.
- It does support IPv6 and requires IPv6 Unicast Global address on the interface of tunnels.
- NHRP signals IPsec tunnel creation and tear down of tunnel, and vice versa IPsec signals NHRP while encryption is lost or cleared.
- NHRP has routing functionality but still routing protocol to be used with DMVPN.
- While no routing protocol is used between spokes, routing table is updated with NHRP "routes".
- In Hub Redundancy model, IPsec failover is stateful but NHRP failover is not stateful.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

- It is possible to enforce QoS Policies per NHRP group \ per tunnel on the Hub.
- NHRP limitations in 2nd phase DMVPN are addressed in 3rd phase DMVPN by having good routing alignment between NHRP routes and Routing protocol (RP) routes.

Commands Used in Simulation on Virtual Interface Tunnel0:

```
ip nhrp authentication    *******
ip nhrp map multicast dynamic
ip nhrp network-id 1
```

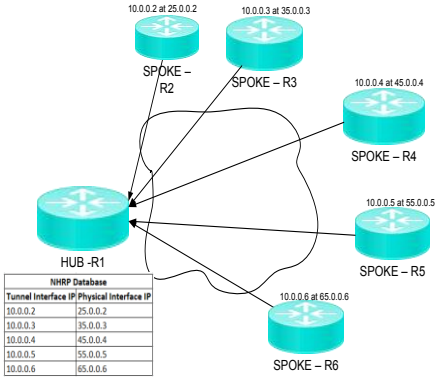


Fig 5. NHRP Registration Process

Configuration of NHRP Spokes does include the NHRP Hub IP address. During the new spoke inclusion in the network, it relays its (spoke's) Physical IP Address (which is bound to physical interface) and Logical IP Address (which is bound to Tunnel Interface) of the establishing tunnel to the Hub.

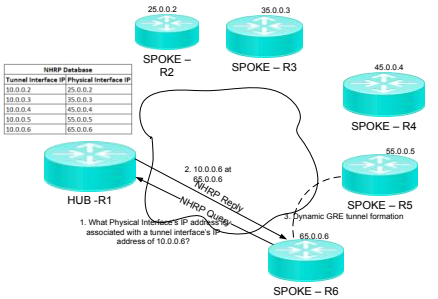


Fig 6. NHRP-Query-Response

In this simulation, while Spoke R6 router needs to establish tunnel with Spoke R5 Router it does not know the physical IP address that corresponds to the virtual tunnel IP address. These below are the steps in discovering the physical IP address of destination location router.

Step 1: Spoke-R6 Router sends query to the HUB-R1 asking the physical address of Spoke-R5 that is corresponding to tunnel interface IP address 10.0.0.5 .

Step 2: Hub-R1 responds to the query to Spoke-R6 that Physical IP address of the tunnel corresponding to 10.0.0.5 is 55.0.0.5 which is Spoke-R5.

Step 3: Having the IP learned from Hub, Spoke-R6 now established the GRE tunnel to Spoke-R5.

```
R1#sh ip nhrp
10.0.0.2/32 via 10.0.0.2
   Tunnel0 created 01:23:42, expire 01:53:31
   Type: dynamic, Flags: unique registered used
   NBMA address: 25.0.0.2
10.0.0.3/32 via 10.0.0.3
   Tunnel0 created 01:23:41, expire 01:53:31
   Type: dynamic, Flags: unique registered used
   NBMA address: 35.0.0.3
10.0.0.4/32 via 10.0.0.4
   Tunnel0 created 00:06:00, expire 01:54:00
   Type: dynamic, Flags: unique registered used
   NBMA address: 45.0.0.4
10.0.0.5/32 via 10.0.0.5
   Tunnel0 created 00:03:20, expire 01:56:40
   Type: dynamic, Flags: unique registered used
   NBMA address: 55.0.0.5
10.0.0.6/32 via 10.0.0.6
   Tunnel0 created 00:11:50, expire 01:54:44
   Type: dynamic, Flags: unique registered used
   NBMA address: 65.0.0.6
```

Fig 7. NHRP detail (output of command "show ipnhrp" on Hub location router in simulation.)

3.3. Internet Protocol Security (IPsec):

IPsec is extensively utilized network layer security mechanisms in the enterprise network. It ensures secure communications across a LAN, WANs, and across the Internet. It provides security by using two different set of rules, AH (Authentication Header) andESP (Encapsulation Security Payload). AH ensures authentication and integrity of data,while the ESP protocol provides encryption, integrity an optional authentication to the data. The IPsec protocol activates in two approaches; tunnel approach and transport approach, the tunnel approach restores the novel IP header and summarizes the complete packet, where as in second approach does not alter the original header and inserted in OSI model among the network layer and the transport layer.

Table: 3. IPSec parameters used in simulation.

IKE PhaseIEncryption and Hash	IKE PhaseII Encryption and Hash	Auth	DH	Tunnel Mode
3DES and SHA	3DES and MD5	Pre-share	GroupII	Transport

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

In Transport mode, it uses Orginal IP Header instead of adding an additional tunnel header. This works well where there is an issue if packet size is increased. Transport mode is generally used in Remote VPN connections where a desktop \ laptop is connected to VPN Concentrator situated in Organizations Data enter i.e., VPN client software installed on user machine connects to VPN server.

Table 4. Transport Mode

ESP Auth	ESP Trailer	Payload	ESP Header	Original IP Header

In Tunnel mode, it uses New IP Header by encapsulating the original packet and has new Source IP and Destination IP. This is majorly used in B2B VPN connections where Source and Destination IP s are VPN peer IPs of either side. Also we have used the same in our Simulation.

Table 5. Tunnel Mode

ESP Auth	ESP Trailer	Payload	Original IP Header	ESP Header	New IP Header

Primary Steps involved establishing Site to Site VPN Tunnel**:**

1. Data is initiated from one router which would be the interesting traffic.

2. Either side routers \ VPN devices would negotiate security association to form IKE phase1 tunnel (ISAKMP tunnel).

3. Within the ISAKMP tunnel protection, IKE Phase II tunnel is negotiated and formed. This IKE Phase II tunnel is called IPSec tunnel.

4. Within this IKE Phase II (IPSec tunnel), the traffic can be restricted only to particular source and destination IP subnets through a ACL (Access Control List). So that non interesting traffic is sent outside the IPSec tunnel protection.

5. After transmission of data over the tunnel, if no intersting traffic is flown for certain time, the tunnel SA is deleted and tunnel is torn down.

In our simulation of DM-VPN, the IKE phase 1 or ISAKMP status is as below described.

QM_IDLE is the state which confirms that SA (Security-Association) is formed and it is formed only if either side parameters are matched. Other statesis MM_KEY_EXCH where it conveys that key is not matched or IP addresses are differently configured on either side.

```
R1#sh cryp isakmp sa
IPv4 Crypto ISAKMP SA
dst             src             state          conn-id status
15.0.0.1        35.0.0.3        QM_IDLE           1002 ACTIVE
15.0.0.1        55.0.0.5        QM_IDLE           1005 ACTIVE
15.0.0.1        45.0.0.4        QM_IDLE           1004 ACTIVE
15.0.0.1        65.0.0.6        QM_IDLE           1003 ACTIVE
15.0.0.1        25.0.0.2        QM_IDLE           1001 ACTIVE
```

Figure: 3.3.1. IKE Phase I Security-associations table at Hub-R1 router.

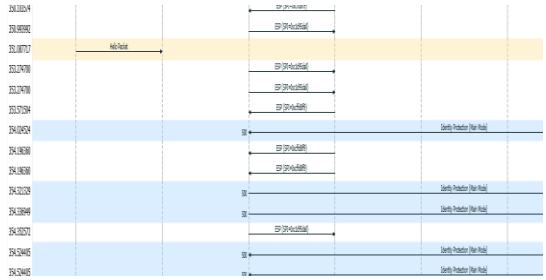


Fig 8. Packets sniffed in Wireshark tool while ISAKMP tunnel is being established.

To understand if our VPN is working correctly from either side, we use command “Show Crypto Engine Connections Active” where it displays the counters of encrypted and decrypted packets. If IKE Phase II or IPSec configuration of any side is configured incorrectly, then we may observe that packets are encrypting but response packets are not decrypted and it may be vice versa. This count should be matching to number of packets sent from other end VPN device as encrypted to our decrypted packets count.

```
R6#sh crypto engine connections active
Crypto Engine Connections

  ID  Type   Algorithm          Encrypt  Decrypt LastSeqN IP-Address
   1  IPsec  3DES+MD5                0        4        4 65.0.0.6
   2  IPsec  3DES+MD5                6        0        0 65.0.0.6
   3  IPsec  3DES+MD5                0      392      392 65.0.0.6
   4  IPsec  3DES+MD5              395        0        0 65.0.0.6
1001  IKE    SHA+3DES                0        0        0 65.0.0.6
```

Fig 9. Secure Connection table in Spoke-R6 router.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
Crypto Engine Connections

     ID  Type    Algorithm           Encrypt  Decrypt LastSeqN IP-Address
      7  IPsec   3DES+MD5                  0      570      570 15.0.0.1
      8  IPsec   3DES+MD5                565        0        0 15.0.0.1
      9  IPsec   3DES+MD5                  0      547      547 15.0.0.1
     10  IPsec   3DES+MD5                544        0        0 15.0.0.1
     11  IPsec   3DES+MD5                  0        6        6 15.0.0.1
     12  IPsec   3DES+MD5                  4        0        0 15.0.0.1
     13  IPsec   3DES+MD5                  0      370      370 15.0.0.1
     14  IPsec   3DES+MD5                368        0        0 15.0.0.1
     15  IPsec   3DES+MD5                  0      285      285 15.0.0.1
     16  IPsec   3DES+MD5                281        0        0 15.0.0.1
     17  IPsec   3DES+MD5                  0        2        2 15.0.0.1
     18  IPsec   3DES+MD5                  1        0        0 15.0.0.1
     19  IPsec   3DES+MD5                  0      245      245 15.0.0.1
     20  IPsec   3DES+MD5                243        0        0 15.0.0.1
   1001  IKE     SHA+3DES                  0        0        0 15.0.0.1
   1002  IKE     SHA+3DES                  0        0        0 15.0.0.1
   1003  IKE     SHA+3DES                  0        0        0 15.0.0.1
   1004  IKE     SHA+3DES                  0        0        0 15.0.0.1
   1005  IKE     SHA+3DES                  0        0        0 15.0.0.1
```

Fig 10. Secure Connection table in Hub-R1 router.

Below output shows the IKE Phase II or IPSec tunnel status on one the Router. It displays the parameters we used in establishing the tunnel and the Peer IP addresses involved with the data of encrypted and decrypted packets. This command "SHOW CRYPTO IPSEC SA" is very helpful in troubleshooting and to understand the parameters mismatch in case encryption is happening from only one side and decryption on other side.

```
R1#show crypto ipsec sa

interface: Tunnel0
    Crypto map tag: Tunnel0-head-0, local addr 15.0.0.1

   protected vrf: (none)
   local  ident (addr/mask/prot/port): (15.0.0.1/255.255.255.255/47/0)
   remote ident (addr/mask/prot/port): (25.0.0.2/255.255.255.255/47/0)
   current_peer 25.0.0.2 port 500
     PERMIT, flags={origin_is_acl,}
    #pkts encaps: 382, #pkts encrypt: 382, #pkts digest: 382
    #pkts decaps: 349, #pkts decrypt: 349, #pkts verify: 349
    #pkts compressed: 0, #pkts decompressed: 0
    #pkts not compressed: 0, #pkts compr. failed: 0
    #pkts not decompressed: 0, #pkts decompress failed: 0
    #send errors 0, #recv errors 0

     local crypto endpt.: 15.0.0.1, remote crypto endpt.: 25.0.0.2
     path mtu 1500, ip mtu 1500, ip mtu idb (none)
     current outbound spi: 0xAEFB5556(2935903574)
     PFS (Y/N): N, DH group: none

     inbound esp sas:
      spi: 0x246331AA(610480554)
        transform: esp-3des esp-md5-hmac ,
        in use settings ={Tunnel, }
        conn id: 1, flow_id: 1, sibling_flags 80000040, crypto map: Tunnel0-head-0
        sa timing: remaining key lifetime (k/sec): (4265861/2167)
        IV size: 8 bytes
        replay detection support: Y
        Status: ACTIVE(ACTIVE)

     inbound ah sas:

     inbound pcp sas:

     outbound esp sas:
      spi: 0xAEFB5556(2935903574)
        transform: esp-3des esp-md5-hmac ,
        in use settings ={Tunnel, }
        conn id: 2, flow_id: 2, sibling_flags 80000040, crypto map: Tunnel0-head-0
        sa timing: remaining key lifetime (k/sec): (4265855/2167)
        IV size: 8 bytes
        replay detection support: Y
        Status: ACTIVE(ACTIVE)

     outbound ah sas:

     outbound pcp sas:
```

Fig 11. Output from simulation for one Crypto IPSec tunnel.

IPSec Commands used in simulation:

```
!
crypto isakmp policy 10
 encr 3des
 authentication pre-share
 group 2
crypto isakmp key ******** address  x.x.x.x
!
!
crypto ipsec transform-set cisco123 esp-3des esp-md5-hmac
 mode tunnel
!
!
crypto ipsec profile dmvpnprof
 set transform-set cisco123
!
```

Other Commands used in troubleshooting are as below:

Pre debug:
Logging
No logging
Service timestamps debug datetime msec
Service timestamps log datetime msec'

During Debug:
Debug crypto isakmp
Debug crypto ipsec
Debug Crypto engine
Undebug all

3.4. Routing protocols:

The exchange of routing tables between the Hub and spoke of the DMVPN is responsible by the dynamic Routing protocols. The routers analyse data and finds an optimal route for data transmission. DMVPN design utilized the different routing protocols, those are EIGRP, OSPF.

- While NHRP has routing functionality, DMVPN still need routing protocol.
- Theoretically all IP based routing protocols are supported.
- HUBs and Spokes Routing
  - HUBs receive network routes from Spokes while using Routing Protocols.
  - Spokes local networks are advertised to other spokes through Routing protocol.
  - I & III Phase have feasibility to use summary routes (Exception OSPF)
  - No summarization possible in II Phase (limitation of 2Hubs per layout while OSPF is used)
- HUB – HUB Routing
  - I Phase: HUB-HUB and HUB-Spoke can have different Routing Protocols as well as different interfaces.
  - II Phase: HUB-HUB and HUB-Spoke must have same tunnel interface and same routing protocol.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

- III Phase: HUB-HUB and HUB-Spoke can have different Routing Protocoland different tunnel interface.

- Routing between Spokes
  - NHRP routes are added to routing table and no routing protocol is used.

3.4.1. EIGRP (Enhanced Interior Gateway Protocol):

This is an Interior Gateway Protocol and improved distance vector protocol using DUAL algorithm to analyse the shortest path to a destination in a network.It is Cisco Proprietary protocol. It has very less usage of network resources and stable network utilise only hello packets. Only routing changes are transmitted across while a change occurs on the network, reducing the load on the network. It has path called feasible successor which is having distance less than the feasible distance (current best distance) and it is used as backup if existing best path fails. We use 'no auto-summary' in EIGRP configuration, to enable the use of smaller subnets instead of classful subnets.

EIGRP calculates the best route using metric formula as below shown. Least-bandwidth is the least bandwidth on the path to destination in kbps units i.e., for 10Mbps it is 10^4 Kbps. It can be set manually by command "bandwidth" on the interface of the link connected. Also Sum of delays is the addition of all delays on the path to destination and it can be manually set as well using "Delay" command on the interface. This can be verified by command "Show Interfaces *type number*".

EIGRP Metric calculation formula:

Metric=((10^7 / least-bandwidth) + Sum Of Delays) * 256

EIGRP in DMVPN:

- Feasible successors - distance vector style. It is highly suitable with DMVPN.
- High scalability is possible.
- Fast convergence by 5sec (HELLO) and 15sec (HOLD).
- Filtering capabilities and route tagging supports in controlling the good metric control.
- Load balancing is possible by utilising the feature of EQUAL COST Multipath and Add path support.
- Load balancing across unequal paths is possible from same spoke to multiple Hubs.

Table 3. Configuration of Enhanced Interior Gateway Protocol used in Simulation

HUB	SPOKE
! router eigrp 100 network 192.168.1.0 255.255.255.0 network 10.0.0.1 255.0.0.0 no auto-summary ! interface channel0 no ip split-horizon eigrp 100 no ip next-hop-self eigrp 100 !	! router eigrp 100 network 192.168.1.0 255.255.255.0 network 10.0.0.0 255.0.0.0 no auto-summary !

3.4.2. OSPF (Open Shortest Path First):

This is a connection state protocol which is developed by internet community as a public standard to be used as Non-proprietary protocol. It has developments for instance, validation of routing updates, Variable Length Subnet Masks, information of routes etc., It does support IPv4 and IPv6 both. It contains area 0 which is core area through which there exists a connection to every other area directly or at least virtual connection. This connection would enable other OSPF router areas to learn \ exchange routes between them. Also, there exists Authorized Router and endorsement Authorized Router which are selected, so to have route tables are learned from the same.

DMVPN using OSPF

- It is Link-State protocol and not suitable to DMVPN.
- Problems with Area:
  - It may not be stable while we use Area0 for Hub and Spoke.
  - In general, it is a single area (Non-Area 0) for the whole network in DMVPN.
  - It would increase complexity if multiple OSPF areas are used.
- Metric control would be difficult.
- Spokes received routes through advertised routes are not summarized.
- DMVPN does not scale well while OSPF is used.
- Equal Cost Multi Path (ECMP) Routing is possible and would be good.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 4: Configuration of OSPF used in Simulation.

HUB	SPOKE
! router ospf 10 net 10.0.0.0 0.0.0.255 area 0 net 192.168.1.0 0.0.0.255 area 0 ! interface tunnel0 ipospf network broadcast !	! router ospf 10 net 10.0.0.0 0.0.0.255 area 0 net 192.168.2.0 0.0.0.255 area 0 ! interface tunnel0 ipospf network broadcast ipospf priority 1 !

4. DMVPN Implementation and analysis:

This section presents the implementation of DMVPN with hub and spoke model and created three different scenarios to evaluate the performance of DMVPN. The network design and simulation is carried out using GNS3 software (Graphical network software), which is an open source software, and Cisco IOS (CIOS) images for running on diverse routers, it allows simulation of complex networks. By utilizing GNS3 and CIOS images, we emulate difficult networks with probable results as it is not possible on live environment always. CIOS executes in a virtual location on a PC's and laptop. The graphical front end of GNS3isDynagentoexecute on apex of Dynamics which is the hub program that makes CIOS need possible and provides user-friendly text-based interface.

Here we have configured for 6 / 9 /13 sites network in which one Main Router acts as Hub and captured the statistics with different routing protocols such as EIGRP, OSPF.

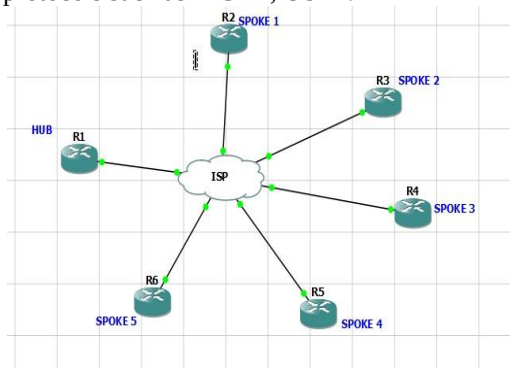


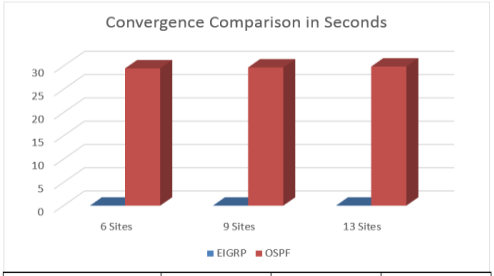
Fig 12: Diagram of Sites connected (1 Hub and 5 Spoke)

Routing Scalability:

As observed with the varied number of sites in DMVPN implementation, the routing protocol EIGRP scales well and has good stability while any new site is appended or disconnected from the organisation network. OSPF has no stability with the increment of sites and with slow convergence, area restrictions; it is not suggested for use in DMVPN of the organisation. EIGRP is more suitable within the organisation and can include BGP while organization mergers \ acquisitions. With the current availability of hardware, the scalability is possible with EIGRP for any organisation while using DMVPN for high resilient and secured network.

Convergence Duration:

Convergence time is low in EIGRP and BGP, while in OSPF it is bit higher. It is only partial routing updates that are being sent in EIGRP while in OSPF it has selection of DR and BDR to process LSAs and DR router relies on other routers in similar area to have similar topology analysis of network in to facilitate area. And thus, there exists a delay in having routing updates slow in OSPF compared to other two EIGRP and BGP. In BGP, it is already defined adjacencies and has slow updates in comparison to EIGRP. Comparing the Internal routing protocols, it is as below:

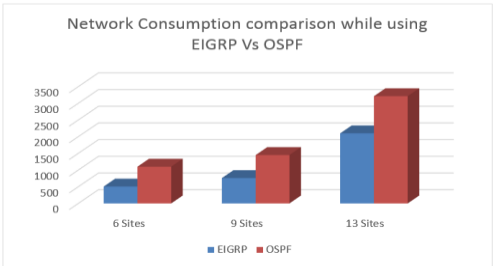


Routing Protocols	6 Sites	9 Sites	13 Sites
EIGRP	0.004	0.006	0.008
OSPF	29.5	29.7	29.9

Fig 13: Convergence statistics comparing EIGRP versus OSPF

Traffic Sent:

Traffic sent is low in EIGRP as it sends only routing updates while there is a topology change in the network. OSPF does full link state database updates and thus send higher size as compared to EIGRP. While comparing the 2 protocols, it is only EIGRP which is sending low data and not utilizing more than half of the link bandwidth.



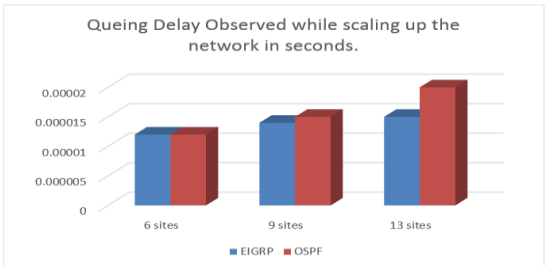
Routing Protocols	6 Sites	9 Sites	13 Sites
EIGRP	505.234	760.365	2100
OSPF	1101	1450	3214.23

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fig 14 : Network Consumption statistics comparing EIGRP versus OSPF (in number of bits)

Queuing Delay:

Queuing delay is measurement of packet wait time before being transmitted outside. This may vary depending on router configuration and network load. This also an important factor while scaling up the network and choosing the right routing protocol is important to decrease the load on router and links, thereby reducing the Queued packets count at that point of time.



Routing Protocols	6 sites	9 sites	13 sites
EIGRP	0.000012	0.000014	0.000015
OSPF	0.000012	0.000015	0.00002

Fig 15. Queing Delay observed while scaling up the network in seconds.

5. CONCLUSION:

In this paper, we studied and worked through the DMVPN with varied number of sites with different routing protocols EIGRP, OSPF. As observed with the parameters such as convergence, consumption and queuing delay, it is observed that EIGRP which is more suitable when in comparing to other routing protocols in terms of configuring and maintenance of the secured network. OSPF is not scaling well and would be complex as areas are restricted in DMVPN. It is suggested to have EIGRP implemented in larger network environments and suggested to use only in small network environments where network vendor is different from Cisco.

6. FUTURE ENHANCEMENT:

In future, we would like to analyse DMVPN performance while using CoSand QOS for VoIP and other real-time applications in bandwidth limited networks. Real-time applications does include Voice, Video, High Sensitive Financial transactions etc.,

7. REFERENCES:

[1]Dynamic Multipoint IPsec VPNs (Using Multipoint GRE/NHRP to Scale IPsec VPNs), http://www.cisco.com/c/en/us/support/docs/security-vpn/ips negotiation-ikeprotocols/41940-dmvpn.pdf

[2]Hanks, Stan, David Meyer, Dino Farinacci, and Paul Traina.RFC 2784-"Generic routing encapsulation (GRE)." (2000).

[3]J. Luciani, D. Katz, D. Piscitello, B. Cole, and N. Dora swamy, “NBMA Next Hop Resolution Protocol (NHRP),” RFC Editor, RFC2332, Apr. 1998.

[4]Cisco Systems, “Developmental Phases of DMVPN and NHRP,” in *NHRP*, Cisco Press, 2007, pp. 6–8.

[5]Kent, Stephen. IP authentication header. RFC 4302, December, 2005

[6]R. White, J. Ng, D. Slice, S. Moore, and others, “Enhanced Interior Gateway Routing Protocol,” 2014.

[7] Moy, J.; OSPF Version 2,The Internet Society. OSPFv2,1998.

[8]N. Angelescu, D.C. Puchianu, G. Predusca, L.D. Circiumarescu, G. Movila, “DMVPN simulation in GNS3 network simulation software.”Electronics, Computers and Artificial Intelligence 2017 - International Conference – 9th Edition

[9]A. Bahnasse, N. El Kamoun, “Study and analysis of a Dinamic Routing Protocols’ scalability over a Dynamic Multi-point Virtual Private Network”, International Journal of Computer Applications, volume 123, no.2, august 2015, pp.26-31.

[10]R. Jankuniene and I. Jankunaite, “Route creation influence on DMVPN QoS”, Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces, Dubrovnik, pp. 609-614, 2009

[11]F. Bensalah, N. El Kamoun, A. Bahnasse, “Analytical performance and evaluation of the scalability of layer 3 tunneling protocols: case of voice traffic over IP”, IJCNS International Journal of Computer Science and Network Security, vol. 17, no.4, April 2017, pp.361-369.

[12] GNS3 Documentation (21, March 14).[Online]. Available: http:// www.gns3.net

[13] Chen, H. (2011, May). “Design and implementation of secure enterprise network based on DMVPN. In Business Management and Electronic Information (BMEI)”, 2011 International Conference on (Vol. 1, pp. 506-511(pp. 1842-1845). IET

[14] Ahmad Karim, Minhaj Ahmad Khan "Behaviour of Routing Protocols for Medium to Large Scale Networks", Australian Journal of Basic and Applied Sciences, 5(6): 1605-1613, 2011.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

A SCALABLE SHARING OF BIG DATA USING AN EFFICIENT SECURITY MECHANISM FOR PRESERVING PRIVACY

Johny Antony P
Research Scholar
NGM College, Pollachi, Tamilnadu
Email: johnypkt@yahoo.co.in

Dr. Antony Selvadoss Thanamani
Head, Department of Computer Science
NGM College, Pollachi, Tamilnadu
Email: selvadoss@gmail.com

***Abstract*— Developing a secure data storage and retrieval system in cloud is one of the crucial and demanding task due to the increased usage of big data. It has been attracted by many healthcare industries and academic communities. In EHR sharing environment, increased cost, reduced security and poor usability are the major causes, which affects the usage of the storage system. So, the traditional works intend to design various security mechanisms include key generation, distribution, privacy preservation, encryption and decryption for cloud data storage. But, it remains with the limitations of increased computational complexity, time complexity, data incorrectness, and more storage space. In this paper medical data is considered as big data. Thus, this paper intends to introduce a new security mechanism by utilizing a score generation, anonymization and encryption techniques. The sensitive and non-sensitive attributes are separated from the given dataset by using the Privacy Score Generation Algorithm (PSGA). Then, the privacy of patient's medical information and health information are preserved by employing an Enhanced k-Incognito Anonymization Algorithm (EkIAA). Finally, the Attribute based Homomorphic Encryption (ABHE) mechanism is used to encrypt the original data by converting it into cipher format. The experimental evaluation analyze the results of the proposed security system by using various measures such as sensitive score, key generation time, cost consumption, encryption time, decryption time, and error rate. At last, the betterment of the proposed technique is proved by relating it with some other existing security mechanisms.**

***Index Terms*— Electronics Health Record (EHR), Cloud Security, Data Storage, Privacy Score Generation Algorithm, Enhanced k-Incognito Anonymization Algorithm, and Effective Attribute Based Homomorphic Encryption.**

I. INTRODUCTION

CLoud is the next generation Information Technology (IT) that offers different services to its users. Due to its self-service, on-demand access, ubiquitous network access, and location dependent resource pooling, it has been used in many enterprises and applications[1, 2]. The Electronics Health Record(EHR) isone of the widely used health record system, where the providers and patients can able to access the patient's information[3]. Normally, the patient have separate health care physicians, therapists, and specialists[4]. Also, they have various insurances like dental, vision, and medical, so a patient's EHR could be scattered around various health sectors[5]. Moreover, it is defined as the repository of the patient data in a digital form. The major characteristics of EHR [6] management are listed as follows:

1. Medical information and data – It holds a defined dataset for ensuring an improved access on the data.
2. Results management – Typically, this feature is more useful for managing all types of results such as laboratory test reports, radiology procedures and other medical reports with increased efficient and reduced cost.
3. Order management – The workflow processes can be improved, and ambiguities can be reduced based on an illegible script.
4. Assessment – The computerized assessment systems have the capacityto increase the clinical performance based on the facets of rule based alerts and reminders.
5. Communication and connectivity – An active communication was established with high quality health care.
6. Administrative process – The efficiency of health care units is increased by the use of electronic scheduling systems, and also it provides the best services to the users.
7. Population health management –It enables the process of reporting with reduced time consumption and labor intensive.

In order to ensure the reliable data storage and retrieval in cloud, the requirements such as data confidentiality, availability, preservability, and integrity should be provided to the cloud[7]. The major threats that affects the cloud data service are data leakage, illegal access, data corruption, and user privacy breach. These issues must be solved by providing proper solutions to the secure cloud data storage, which includes Searchable Encryption (SE)[8], Homomorphic Encryption (HE)[9], selective encryption[10], Attribute Based Encryption (ABE)[11], Proof of Retrievability (POR)[12], Provable Data Possession (PDP)[13], access pattern protection[14], query privacy protection[15], and identity privacy protection. These mechanisms provide the security to the data search, data computation, data storage, and data access based on privacy preservation.

A. Problem Identification

In the existing works, various security mechanisms have been proposed for ensuring the privacy and confidentiality of the EHR stored in cloud. The cryptographic management techniques are used to solve the security issues by generating the keys used for data access. It allows only the authorized

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

users to access the data by managing the role based and attribute based access control policies. Moreover, the identity based encryption and decryption techniques are used to secure the data storage. Then, the key generation, distribution, signature generation, and verification processes have been performed to ensure the correctness and confidentiality of the data. Still, it remains the following limitations: it requires a fines grained access control and authorization policies, it relies on the fully trusted party, and increased computational complexity. Due to these problems, this paper objects to design an enhanced mechanisms for guaranteeing the security of the EHR stored in cloud.

B. Objectives

The research goals focused on this work are as follows:

- To identify the sensitive and non-sensitive data attributes, the Privacy Score Generation Algorithm (PSGA) is developed.
- To hide the patient's details like medical and personal information, an Enhanced k-Incognito Anonymization Algorithm (EkIAA) is introduced.
- To encrypt the data in an efficient manner, an Attribute Based Homomorphic Encryption (ABHE) technique is implemented.

C. Organization

The remaining sections present in the paper are arranged as follows: the standard techniques and architectures correlated to EHR security are surveyed in Section 2. The working procedure of the proposed methodology is presented with its detailed flow and algorithm illustration in Section 3. The performance results of the existing and proposed techniques are evaluated and compared by using various performance measures in Section 4. At last, the overall conclusion of the paper is presented in Section 5.

II. Related Works

This section surveys various security mechanismassociated to EHR cloud data storage and retrieval.

Wu, et al [16] developed a New Extensive Data Access Control – Multiauthority Cloud Storage System (NEDAC-MACS) for guaranteeing a secure attribute revocation. In this system, two attackers were constructed based on the vulnerabilities of the revocation security. Here, the monotone access structure was defined by the data owner with respect to the data's logic attribute granularities for encrypting the data. During attribute revocation, the key generation, and secret key updation processes were performed. *Wang, et al*[17]intended to ensure the privacy of the cloud data by developing a secure storage system with public auditing mechanism. Here, the integrity of the data was ensured and the cloud computing resources were saved for reducing the online burden. Then, the integrity of the data stored in cloud could be validated by using the TPA. After that, the data outsourced by the use could be validated with the help of public auditing mechanism. Here, multiple auditing tasks were performed in a batch manner by using the batch auditing algorithm. The merit of this work were better efficiency and security, but has the limitation ofhigh computational complexity. *Sookhak, et al* [18] suggested a dynamic Remote Data Auditing (RDA) scheme for securely storing the big data in cloud. In this paper, the Divide and Conquer Table (DCT) was utilized to perform the data operations like insertion, updation, deletion, and append. The major considerations have been mainly focused on this work were listed as follows:

- Efficiency
- Frequency
- Detection probability
- Public/private verifiability
- Dynamic update

Typically, a large number of data blocks have been rebalanced within the data structure, which leads to increased computational cost at the auditor side. The entities involved in this design were data owner, TPA, Cloud Service Provider (CSP) and user. This work does not verify the integrity of the files stored in the distributed systems. *Wang, et al* [19] utilized an identity based cryptographic techniques for developing a secure E-health system. The stages involved in this work were key generation, encryption, re-encryption, and decryption. Moreover, the Identity Based Encryption (IBE) and Identity Based Proxy Re-Encryption (IBPRE) schemes were integrated to improve the confidentiality. Then, the bilinear algorithm was used to estimate the polynomial time during encryption and decryption. *Yuksel, et al* [20]surveyed various techniques used for Electronic Health Service (EHS) based on measures of security, privacy, and integrity. The major contributions have been focused on this work were as follows:

- A method based approach was utilized to systematically analyze the performance of this system.
- The characteristics, components and challenges of e-health services have been investigated.

Also, a distributed architecture was developed based on the Hierarchical Identity Based Public Key Infrastructure (HIB-PKI), which ensured the security of the data storage. Also, the trust management scheme was utilized to establish the trust between the parties in the system. Also, the role based access control mechanism was employed to restrict the limited access on the data. *Feng, et al* [21] developed a secure and efficient dynamic auditing protocol for ensuring the integrity of the data. The intention of this work was to design an auditing framework for increasing both the privacy and efficiency of the storage system. Also, a load distribution strategy was utilized to reduce the computational overhead at the client side. From the work, it was analyzed that the suggested mechanism has the ability to handle the errors and reduce the overhead in an efficient manner. However, it does not use any verification scheme for proving the efficacy of the suggested technique. *Qian, et al* [22] developed an multi-authority based ABE mechanism for ensuring the privacy preservation of the Personal Health Record (PHR). Here, an Attribute Based Encryption (ABE) mechanism was utilized to encrypt the PHRs before storing it into the cloud. Then, an on-demand user/attribute revocation mechanism was employed for revocation and dynamic policy updation. During the security analysis, the collusion resistance, forward secrecy, and data confidentiality have been measured. *Li, et al*[23]developed a

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

patient centric framework for securely accessing the PHR stored in a semi trusted servers. Also, anABEmechanism was utilized to attain a fine grained and scalable access control. In this system, a multiple data owner scenario was considered, which reduced the key management complexity for both users and owners. Based on the user's data access requirements, the system was split into multiple security domains such as personal domain and public domain.

Sahi, et al [24] suggested a disaster recovery plan for securely storing the health records in a third party server. The aim of this paper was to ensure both the privacy and integrity of the EHRs and PHRs. Also, it offered a break glass access feature for recovering the data at an emergent situations. Then, the availability and continuity of the system were guaranteed during the disaster. Moreover, the Parallel Encryption Mode (PEM) algorithm was used to perform the encryption with increased speed. In this technique, each block used a hash value of the shared data for ensuring the data randomness. The entities involved in this design were as follows: data consumers, data owner, trusted party, controller, patients and cloud. The limitation behind this work was increased time consumption and complexity. *Hong, et al* [25] designed a Time and Attribute Factors Combined Access Control (TFAC) mechanism for processing the time sensitive data stored in the cloud. In this framework, the time released encryption technique was embedded with the Cipher-text Policy Attribute Based Encryption (CP-ABE) technique. It categorized different users and offered a fine grained access control. Here, the data owner could autonomously designate the intended users and their access privileges. Moreover, the time sensitive data was outsourced by building a scalable and fine grained access control mechanism. The disadvantage of this work was, it does not prove the superiority of the suggested security technique.

Zhang, et al [26]introduced a Privacy Preserving High Order Possibilistic C-Means (PPHOPCM) algorithm for protecting the privacy of the big data stored in cloud. In which, the membership matrix was updated and the clustering centers were approximated for supporting the secure computing. Also, an anonymous key issuing protocol and enforce write access control mechanism have been utilized to ensure the security of the data. *Lou, et al*[27]developed a Mining Associations with Secrecy Konstraints (MASK) technique for securing the data stored in cloud by forming the privacy preservation rules. Two different strategies such as data perturbation and query restriction were utilized to establish the privacy preservation. Then, the DPQR technique was used to process the original data in different way with reduced time complexity. The advantage of this paper was, it attained a better execution efficiency by using the MASK algorithm. However, it was only suitable for the Boolean data, and it does not work for the numerical and other types of data. *Wang, et al[28]*developed a Privacy Preserving Single Layer Perceptron (PSLP) model for processing the big data in cloud. The major intentions of this work were to reduce the computational cost and communication overhead. Then, the paillier cryptosystem was utilized to attain the Homomorphic properties in a privacy preserving applications. This system includes the stages of key generation, encryption and decryption. Moreover, the Single Layer Perceptron (SLP) scheme was employed for classification that integrates the system setting and privacy preservation policies.

Wei, et al [29]utilized a combination of Principle Component Analysis (PCA) and random matrix theory techniques for analyzing the privacy preservation of the cloud data. A sample complexity bound was determined by defining the number of principal components. Moreover, the empirical simulations have been conducted to demonstrate the performance of the suggested system. *Yuan, et al*[30]suggested a k-means clustering technique for processing the large scale dataset with better privacy preservation. Here, the MapReduce framework was utilized to ensure the security and efficiency of the data storage. In this paper, two different threat models have been considered that includes ciphertext only model and known background model. Moreover, the k-means clustering technique was employed to reallocate a set of data objects into *k* number clusters. Then, the data objects and centers were represented as the multi-dimensional vectors, and its distance was estimated by using the Euclidean distance measure. The benefits of this work were improved scalability, efficiency, and accurateness. Still, it has an increased computational complexity, which must be reduced by simplifying the process of privacy preservation.

Zhu, et al[31]formed a pre-diagnosis framework by using a non-linear Support Vector Machine (SVM) technique for increasing the security of medical information. In this system, the authorized data analysis organization was considered to offer an online medical pre-diagnosis service for the listed medical users. By using the decision classifier, the service provider estimates the encrypted medical data. Here, the query vector was used to represent the medical information, in which the listed users could request the services. The major security requirements have been considered in this work were privacy, authentication, and confidentiality. The limitation of this work was, it required to demonstrate the efficiency of the security system.*Zhang, et al* [32] implemented a proximity aware clustering technique for processing the big data. Here, a Single Objective Proximity Aware Clustering (SPAC) technique was utilized for performing the data anonymization. Then, the proximity aware distance measure was utilized for detemining the distance between the clusters. The data parallel computation was performed for increasing the time efficiency. However, it required to increase the scalability and availability of the big data.

From the survey, it is analyzed that the existing security mechanisms and frameworks have both its own merits and demerits. The limitations studied are as follows:

- It does not ensure the correctness of the data
- Increased computation and storage overhead
- It relies on the full trust of third party
- Lot of storage in private cloud

Thus, this paper intended to introduce a new security mechanism for the EHRs stored in cloud.

III. Secure EHR Data Storage System

This segment presents working procedure of the proposed methodology with its algorithmic and flow illustration. This paper intends to design a secure cloud data storage system by implementing an enhanced security algorithms. The working

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

flow of the proposed system is depicted in Fig 1, in which the medical data and personal information about the patients are separated from the given dataset at the initial stage. Then, the hospital admin can generate the privacy score for splitting the data into sensitive and non-sensitive attributes by using the Privacy Score Generation Algorithm (PSGA). After that, an Enhanced k-Incognito Anonymization (EkIA) technique is implemented for ensuring the privacy preservation. Consequently, the original patient's data are encrypted by the use of an effective Attribute based Homomorphic Encryption (ABHE) mechanism, which efficiently increases the security of data stored in a cloud. Finally, the encrypted data can be forwarded to the user, who give the request to the service provider. The stages involved in this system are described as follows:

- Privacy score generation
- Privacy preservation
- Attribute based encryption

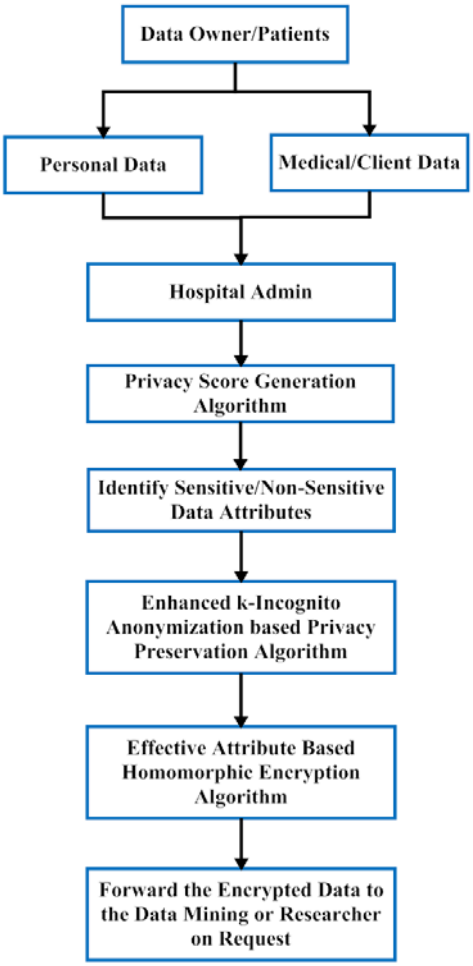


Fig 1. Flow of the proposed system

A. Privacy score generation algorithm

The privacy score is one of the indicator that is used to measure the user's potential risk due to the online information sharing. By using the score, the cloud user can monitor the privacy risks in real time. Moreover, this score satisfies the properties of more sensitive and more visible for the information revealed by the user. In this work, the Privacy Score Generation Algorithm (PSGA) is utilized to generate the privacy score for splitting the sensitive and non-sensitive attributes from the dataset. In this algorithm, the overall medical report and individual patient's report are taken as the input for generating the privacy score. For each record in the medical report, the abnormal and normal classes are identified by comparing it with the patient's individual report. Based on the class, the data counting value (i.e. calculation variable) can be updated. Then, the sum variables can be initialized with the values of calculation variables separately. Consequently, the sensitivity matrix and likert scale matrix are updated, then the visibility matrix is formed based on the individual's score calculation. Similarly, the minimum index matrix is constructed for finding the minimum score value stored in the matrix. After that, the sensitivity score (i.e. β matrix) is formed based on the attribute score calculation. By using this value, the privacy score is generated for separating the sensitive and non-sensitive attributes. The working procedure of the proposed PSGA is illustrated as follows:

Algorithm I – Privacy Score Generation Algorithm

Input: *Data (M_d), Overall Patients medical report (M_r)*
Output: *Privacy score matrix (PR (i, j))*
Procedure:
Let (M_r) is Overall Medical Report.
Let (U_r) is Patient's individual reports.
Take (M_r) and (U_r) data's.
// Privacy Score Calculation
Step 1: $(M_r) \leftarrow Medical\ Data$
Step 2: *for x=1 to n (n←* $M_r length$*)*
Step 3: *for y=1 to n1 (n1←* $M_r\ length$*)*
Step 4: *if (*$U_r \leq (M_r.equals()_{abnormal})$*)*
Step 5: *C_value =1;*
Step 6: *else if (*$U_r \leq (M_r.equals()_{normal})$*)*
C_value =0;
Step 8: Sm_i*=0,* $Sm_j = 0$;
Step 9: $Sm_i = value$*←(C_value =1),* $Sm_j = value$*←(C_value=0)*
Step10: $Sen\_max_{[i][j]} \leftarrow likert\_matrix_{[smi][smj]}$
Step 11: *else* Sm_j *→generate sensitive score (* $Sen\_max_{[i][j]}$*)*
Step 12: *End* Sm_i, Sm_j
Step 13: *Calculate* $V\_matrix_{(i,j)}$
Step 14: *for x1=1≤* $Sm_i.len.$
Step 15: *for x2=1 ≤* $Sm_j.len$
Step 16: $M\_index[m_i][m_j]$ *= [*Sm_i *val] [*Sm_j *val];*
Step 17: $P_{ci}\{R_1(i,j) = 1\}tt \times 1 init R_1 = n$ *to* $Sm_i.len;$
Step 18: $P_{cj}\{R_2(i,j) = 0\} \times 0 init R_2 = n$ *to* $Sm_i.len;$
Step 19: $V\_matrix_{(i,j)} = P_{ci} \times P_{cj}$ *init* 0
Step 20: *End* $V\_matrix_{(i,j)}$
Step 21: *find* $score(Math\_index[mi][mj])$
Step 22: *Calculate Sensitivity Score* βi
Step 23: $R_i = \sum_n k = k \sum_{n-1}^{0} Math\_index_{[mi]} = \pm Math\_index_{[mj]}$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Step 24: $\beta i = \sum_{n=0} \frac{R_i}{M_{i.lenght}} init\ n=0;$
Step 25: $PS=\sum_{n}^{i=0} \beta i \times V(Math\_index[m_i][m_j]$
Step 26: *End* βi

B. Enhanced k-Incognito Anonymization based Privacy Preservation Algorithm

Preserving the privacy of the user, identity and data in cloud plays an essential role, which is used anonymize the data for satisfying the given privacy model. Similarly, the data anonymization has been mainly used to ensure the privacy of the user's data. In which, the user identity and sensitive data are hidden, so the privacy of the user can be effectively preserved. Due to the properties of volume, velocity, and variety, performing the data anonymization with increased scalability and efficiency are the major issues. Also, it enabled the secure sharing of data by enabling the data intactness and retrievability. When the user access the data from the cloud, it protects the multiple dimensions of the privacy information. In this work, the k-incognito anonymization model is mainly used for protecting the privacy of the user's data. In this stage, the outputs obtained from the previous stages such as sensitive(S_{D1})and non-sensitive data (S_{D2}) are taken as the input for anonymization. Here, the root node matrix is constructed with respect to the number of nodes and number of edges. If the root length is greater than the function matrix, the length is updated as 0 and 1. If the match index value is equal to the summation of special character and functional matrix, the match index is inserted into the queue. If the match index is equal to 0, the queue is deleted from the index; otherwise, the index value can be replaced with the support matrix. If the queue is not empty, the index of the queue can be filled with the function matrix. Based on this process, the anonymization is performed for preserving the privacy of the user's data. The working procedure of the proposed EkIA technique is explained as follows:

Algorithm II – Enhanced k-Incognito Anonymization based Privacy Preservation

Input: *Sensitive data* (S_{D1}) *and non-sensitive data* (S_{D2});
Output: *Anonymized* $Aj_{[i][j]}data = (S_{D1}) \times (S_{D2})$ *matrix result;*
Let consider, $(S_{Dt}) \leftarrow (S_{D1}) \pm (S_{D1})$
(U_r) *- Patient's individual reports*
(M_r) *- Overall Medical Report*
$(S_{p[i][j]}) \leftarrow \{Special\ Character\}$
$(fset\ ) \rightarrow 0;$
$C_{node(i)} \leftarrow (U_r)$ *Number of nodes;*
$E_{node(j)} \leftarrow (U_r)$ *Number of edges;*
$H_{queue} \leftarrow$ *Empty queue;*
Step 1: *for* $i = 1$ *to n do*
Step 2: $C_{mt1_{value}} = 0, C_{mt2_{value}} = 0;$
Step 3: $root_{[k]} = \{c_{node(i)}\} \times \{E_{node(j)}\};$
Step 3: *while queue is not empty do*
Step 4: $node$=*remove all item from queue*
Step 5: $f_{[x][y]=}cf_{[i][j]}\{C_{node(i)} \times C_{node(j)}\}$
Step 6: *if* $f_{[x][y].length} \leq root_{[k].lenght}$ *then*
Step 7: $f_{(i)}\_lent = 0 \,||\, f_{(i+1)}\_lent = 1;$
Step 8: *if* $m\_val_{[i][j]} = \sum_{n=0}^{f(i).length} S_{p[i][j]} \pm f_{[x][y].values}$
Step 9: *then* $insert\ into m\_val_{[i][j]}$ *at* $queue(index_{(i)})$
Step 10: *if* $m\_val_{[i][j]} == 0$ *then*
Step 11: $Delete\ queue(index_{(i)})$ *from* $m\_val_{[i][j]}$
Step 12: *end if* $Replace(index_{(i-1)})\ value\ for\ S_{p[i][j]}$
Step 13: *if* $fset = \{queue(size)\}$
Step 14: *if* $\{queue(size)\}! = empty$ *then*
Step 15: $fill\ queue(index) f_{(i)_{lent}} = f_{(i)lent=lent+1}$
Step 16: *end if*
Step 17: *end if*
Step 18: $end\ node\ insert\ ||\ node\ = empty$

C. Attribute based Homomorphic Encryption

The main reason of using Homomorphic encryption is, it allows the arbitrary computations on the encrypted data without using the secret key. In this scheme, the low degree polynomials can be evaluated on the encrypted data. The fully Homomorphic encryption has the ability to perform all types of operations, where the arbitrary number of additions and multiplications have been performed on the encrypted data. The major benefits of using this technique are, it provides increased computing security, and keeps the secret key as secure. Due to these reasons, an Attribute based Homomorphic Encryption (AHE) mechanism is utilized in this work to attain the security property. It allows to perform the arbitrary computations without knowing the secret key. In this stage, generated key, sensitivity and non-sensitivity matrix are considered as the inputs. Then, the column length is estimated for the anonymization data, based on this the length matrix is constructed. Here, two prime integers are declared and the value of λ is computed by multiplying the length matrix and anonymization data matrix. Consequently, the private key and public key used for encrypting the data can be generated, and the data is encrypted by using these keys. The working procedure of this mechanism is illustrated as follows:

Algorithm III – Effective Attribute based Homomorphic Encryption

Input: $As_{key} \leftarrow Key\ \ genS_{[i][j]} \leftarrow Sense\_matrix$ *and* $Sx_{[i][j]} \leftarrow Non\ Sense\_matix$
Output: $Hpv_{key} \leftarrow$ *Private encrypted data;*
$Hpu_{key} \leftarrow$ *Public encrypted data;*
$enc_{[data]} \leftarrow$ *Encrypted data;*
$dec_{[data]} \leftarrow$ *Decrypted data;*
$ka\_mat_{[i][j]} \leftarrow K \leftarrow$ *Anonymization data;*
$s\_matix_{[i][j]} \leftarrow K \leftarrow$ *Sensitive data;*
$Ns_{[i][j]} \leftarrow$ *Non-sensitive data;*
Step 1: *for i = 1 n do*
Step 2: $C_{len} = ka\_mat_{[i][j]}\ lenC_{len}! = null then$
Step 3: *while (j <=1) to n do*
Step 4: $len_{[i1][j1]} = \sum_{t=0}^{n=1} \{\{ka\_mat_{[i][j]}\} \times 1\} \times \{ka\_mat_{[i][j]}.length\};$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Step 5: $A_i \leftarrow prime\ 1; A_j \leftarrow prime\ 2$ *such that* $A_i \neq A_j$ *then;*
Step 6: $n = A_i * A_j$;
Step 7: *Compute* $\lambda = lcm\ (len_{[i1][j1]}.lenght - 1 \times ka_{mat[i][j]}.length - 1)$
Step 8: *Choose* $e \in ka_{mat[i][j]}\ ka_{mat[i][j]}{}^2, n$ *divide by* e
Step 9: *Public key* $\leftarrow (e, n)$;
Step 10: *Private key* $\leftarrow (p, ka_{mat[i][j]}.length - 1)$;
Step 11: $C = m^e\ mod\ \{n \times s\_matix_{[i][j]} + ka\_mat_{[i][j]}\}$;
Step 12: *end if;*
Step 13: *end if;*
Step 14: *end* $n = 0\ ||\ C = 0$ *do continue* $i = 1, j = 1$; *continue;*
Step 15: *else;*
Step 16: *Break;*

IV. Performance Analysis

In this subdivision, theexperimental results of the proposed technique is evaluated by using various measures such as sensitive score, error rate, encryption time, decryption time, cost analysis and key generation time. To prove the efficiency of this work, some of the existing techniques have been considered in this analysis.

A. Sensitive Score Calculation

Fig 2 shows the sensitive score analysis of the existing andproposed techniques with respect to various attributes such as contact number, e-mail, address, birth date, home town, current town, job details and relationship status. The techniques [27]have been considered in this analysis are Fuzzy C-Means (FCM), Average Variance Extracted (AVE), MASK and Center Scalar Matrix (CSM). From the results, it is evident that the proposed mechanism provides the best score results, when compared to the other techniques.

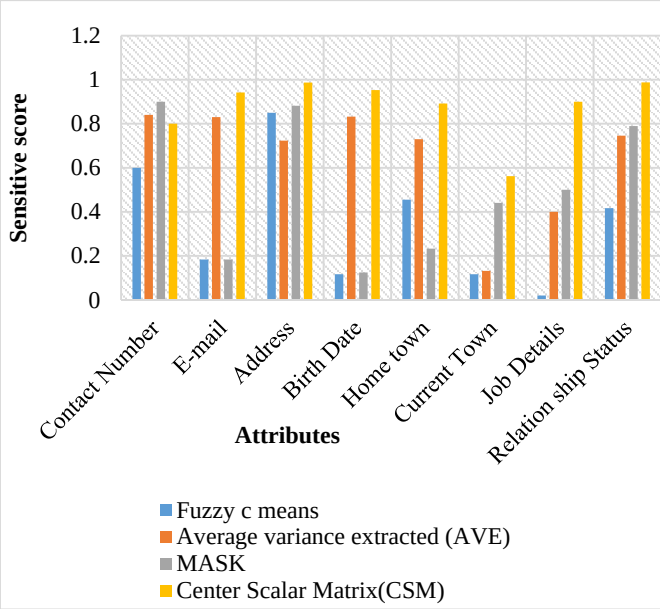


Fig 2. Sensitive score calculation

B. Error Rate

The efficiency of the cloud data storage system is evaluated based on the error rate value, which can be estimated based on the query response. The intention of the cloud data storage system is to reduce the error rate by providing the more relevant results with respect to the user query. Fig 3 evaluates the error rate of both existing [33]and proposed techniques with respect to varying levels. The techniques have been considered in this evaluation are k-anonymity, linear time, Fuzzy C-Means (FCM) [34] and k-incognito anonymization. From the analysis, it is evident that the error rate of the proposed technique is reduced up to 3%, when compared to the existing techniques. Due to the increased scalability and efficiency of the proposed incognito anonymization technique, the error rate is efficiently reduced.

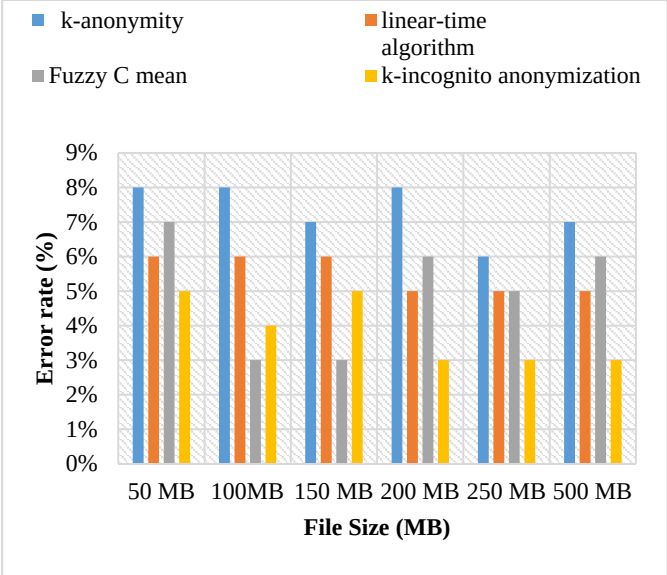


Fig 3. Error rate

C. Runtime of Encryption and Decryption

Encryption time is defined as the amount of time taken for encrypting the original data into a cipher data, which is expressed in terms of milliseconds, and is calculated as follows:

$$ET = End_{time} - Start_{time} \quad (2)$$

Similarly, decryption time is defined as the amount of time taken for decrypting the cipher data into an original data, which is expressed in terms of milliseconds and is calculated as follows:

$$DT = End_{time} - Start_{time} \quad (3)$$

Fig 4 shows the encryption and decryption time of both existing and proposed techniques with respect to varying levels. The techniques have been considered in this analysis are Enhanced and Secured RSA Key Generation Scheme (ESRKGS), hybrid cryptography algorithm, and Dynamic Key Length based Security Framework (DLSeF). From the Fig 4

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

and 5, it is analyzed that the proposed AFHS technique requires reduced time consumption, when compared to the existing techniques.

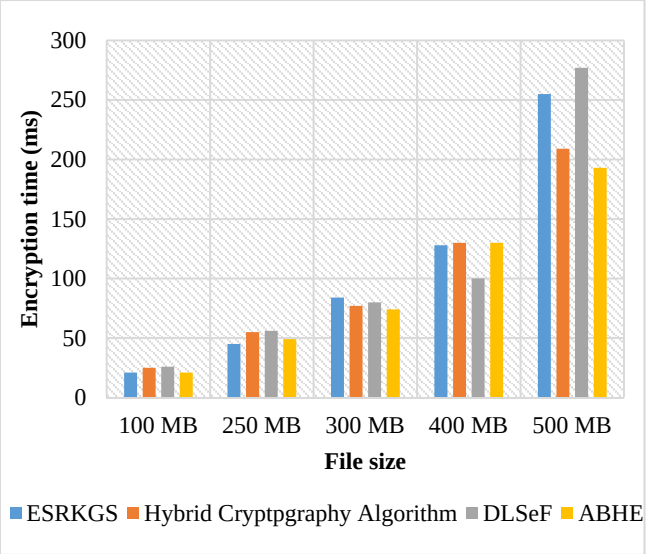


Fig 4. Encryption time

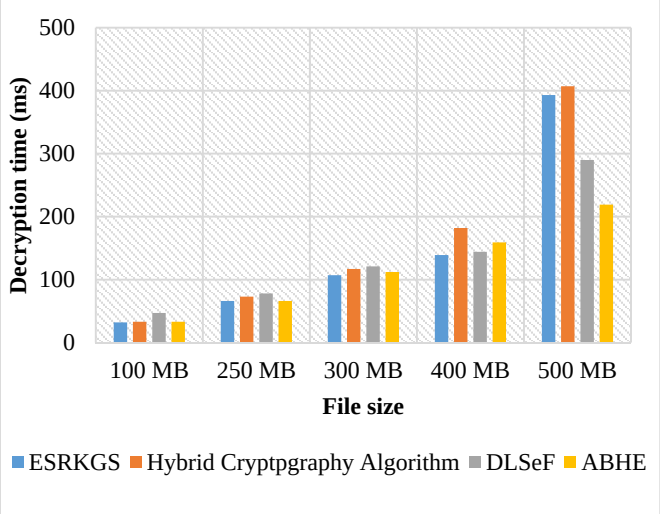


Fig 5. Decryption time

D. Cost Analysis

Fig 6 evaluates the computational cost consumption of the existing and proposed techniques with respect to varying levels. The existing techniques have been considered in this analysis areESRKGS, hybrid cryptography algorithm, and DLSeF. Generally, the encryption technique requires more time and cost consumption for generating the keys during the encryption of the data. In the proposed system, these problems have been solved by implementing an efficient mechanisms. From the results, it is analyzed that the proposed ABHE mechanism requires reduced cost consumption, when compared to the other techniques.

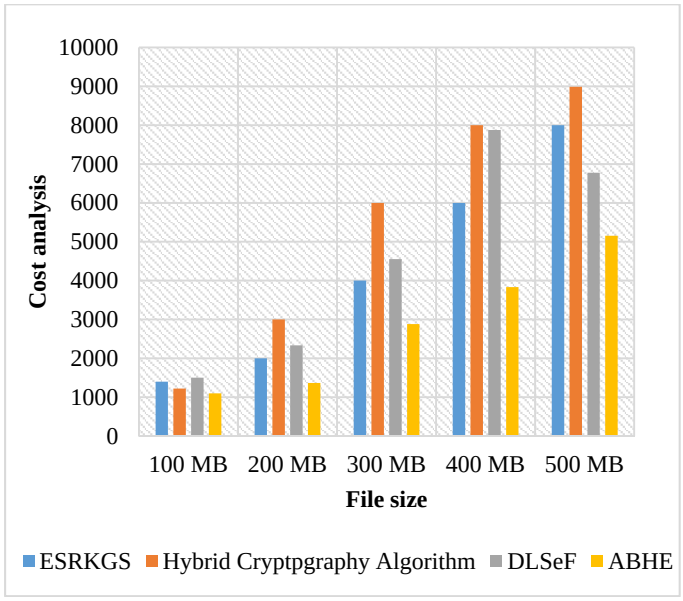


Fig 6. Cost analysis

E. Key Generation Time

Key generation time is defined as the amount of time taken by the encryption technique for generating the key used for data encryption and decryption. Based on this time, the performance of the encryption mechanism can be determined, which is calculated as follows:

$$KT = IT_{time} + E_{time} \quad (4)$$

Where, KT represents the key generation time, IT_{time} indicates the information transferring time, and E_t represents the execution time. In Fig 7, the key generation time of both existing and proposed encryption techniques is evaluated. From the evaluation, it is observed that the key generation time of the proposed ABHE can be efficiently decreased, when compared to the existing techniques.

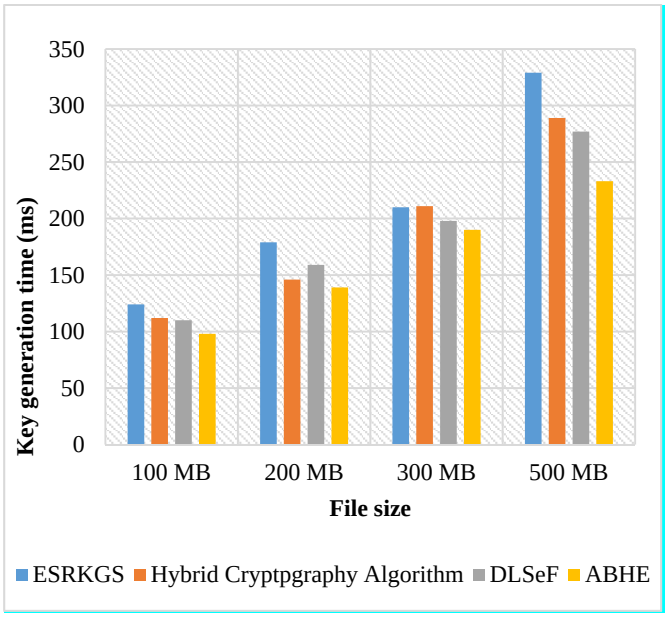


Fig 7. Key generation time

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. Conclusion and Future Work

Providing security to EHR is one of the demanding and essential task in recent days. This work intends to develop a new security mechanism for an EHR storage system. At first, the sensitive and non-sensitive attributes are separated from the dataset by generating the privacy score for the user attributes with the help of PSGA. To hide the patient's personal information and medical information, the data anonymization is performed by using an EkIA algorithm, which also ensured the privacy of the data. Finally, the data encryption is performed by implementing the ABHE mechanism, which generates the cipher data for storage. These security algorithms ensure the secure EHR storage and retrieval, then it provides the benefits like increased efficiency, reduced time and cost consumption. During experimental evaluation, various existing techniques have been considered to prove the efficiency of the proposed technique. For this evaluation, different performance measures such as score generation, cost consumption, key generation time, encryption time, decryption time, and error rate are used. The performance results stated that the proposed EHR security mechanism provides the better outcomes compared than the existing techniques.

REFERENCES

[1] J. Sen, "Security and privacy issues in cloud computing," in *Cloud Technology: Concepts, Methodologies, Tools, and Applications*, ed: IGI Global, 2015, pp. 1585-1630.

[2] M. Ali, *et al.*, "Security in cloud computing: Opportunities and challenges," *Information Sciences,* vol. 305, pp. 357-383, 2015.

[3] U. Premarathne, *et al.*, "Hybrid cryptographic access control for cloud-based EHR systems," *IEEE Cloud Computing,* vol. 3, pp. 58-64, 2016.

[4] A. M. Alshiky, *et al.*, "Attribute-based access control (ABAC) for EHR in fog computing environment," *International Journal on Cloud Computing: Services and Architecture (IJCCSA),* vol. 7, pp. 27-34, 2017.

[5] Z. Cai, *et al.*, "Towards secure and flexible EHR sharing in mobile health cloud under static assumptions," *Cluster Computing,* vol. 20, pp. 2415-2422, 2017.

[6] F. Shahzad, *et al.*, "On the use of CryptDB for securing Electronic Health data in the cloud: A performance study," in *E-health Networking, Application & Services (HealthCom), 2015 17th International Conference on*, 2015, pp. 120-125.

[7] A. Abbas and S. U. Khan, "A review on the state-of-the-art privacy-preserving approaches in the e-health clouds," *IEEE journal of biomedical and health informatics,* vol. 18, pp. 1431-1441, 2014.

[8] Z. Xia, *et al.*, "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data," *IEEE transactions on parallel and distributed systems,* vol. 27, pp. 340-352, 2016.

[9] F. Zhao, *et al.*, "A cloud computing security solution based on fully homomorphic encryption," in *Advanced Communication Technology (ICACT), 2014 16th International Conference on*, 2014, pp. 485-488.

[10] K. Gai, *et al.*, "Privacy-aware adaptive data encryption strategy of big data in cloud computing," in *Cyber Security and Cloud Computing (CSCloud), 2016 IEEE 3rd International Conference on*, 2016, pp. 273-278.

[11] N. S. Kumar, *et al.*, "Enhanced attribute based encryption for cloud computing," *Procedia Computer Science,* vol. 46, pp. 689-696, 2015.

[12] J. Li, *et al.*, "OPoR: enabling proof of retrievability in cloud computing with resource-constrained devices," *IEEE Transactions on Cloud Computing,* vol. 3, pp. 195-205, 2015.

[13] A. F. Barsoum and M. A. Hasan, "Provable multicopy dynamic data possession in cloud computing systems," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 485-497, 2015.

[14] Z. Fu, *et al.*, "Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing," *IEICE Transactions on Communications,* vol. 98, pp. 190-200, 2015.

[15] R. Li, *et al.*, "Fast range query processing with strong privacy protection for cloud computing," *Proceedings of the VLDB Endowment,* vol. 7, pp. 1953-1964, 2014.

[16] X. Wu, *et al.*, "On the security of data access control for multiauthority cloud storage systems," *IEEE Transactions on Services Computing,* vol. 10, pp. 258-272, 2017.

[17] C. Wang, *et al.*, "Privacy-preserving public auditing for secure cloud storage," *IEEE transactions on computers,* vol. 62, pp. 362-375, 2013.

[18] M. Sookhak, *et al.*, "Dynamic remote data auditing for securing big data storage in cloud computing," *Information Sciences,* vol. 380, pp. 101-116, 2017.

[19] X. A. Wang, *et al.*, "Cost-effective secure E-health cloud system using identity based cryptographic techniques," *Future Generation Computer Systems,* vol. 67, pp. 242-254, 2017.

[20] B. Yüksel, *et al.*, "Research issues for privacy and security of electronic health services," *Future Generation Computer Systems,* vol. 68, pp. 1-13, 2017.

[21] B. Feng, *et al.*, "An efficient protocol with bidirectional verification for storage security in cloud computing," *IEEE Access,* vol. 4, pp. 7899-7911, 2016.

[22] H. Qian, *et al.*, "Privacy-preserving personal health record using multi-authority attribute-based encryption with revocation," *International Journal of Information Security,* vol. 14, pp. 487-497, 2015.

[23] M. Li, *et al.*, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," *IEEE transactions on parallel and distributed systems,* vol. 24, pp. 131-143, 2013.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[24] A. Sahi, *et al.*, "Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan," *Computers in biology and medicine,* vol. 78, pp. 1-8, 2016.

[25] J. Hong, *et al.*, "TAFC: Time and attribute factors combined access control for time-sensitive data in public cloud," *IEEE Transactions on Services Computing,* 2017.

[26] Q. Zhang, *et al.*, "PPHOPCM: privacy-preserving high-order possibilistic c-means algorithm for big data clustering with cloud computing," *IEEE Transactions on Big Data,* 2017.

[27] H. Lou, *et al.*, "Data mining for privacy preserving association rules based on improved MASK algorithm," in *Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference on*, 2014, pp. 265-270.

[28] G. Wang, *et al.*, "PSLP: Privacy-preserving single-layer perceptron learning for e-Healthcare," in *Information, Communications and Signal Processing (ICICS), 2015 10th International Conference on*, 2015, pp. 1-5.

[29] L. Wei, *et al.*, "Analysis of a privacy-preserving PCA algorithm using random matrix theory," in *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*, 2016, pp. 1335-1339.

[30] J. Yuan and Y. Tian, "Practical privacy-preserving mapreduce based k-means clustering over large-scale dataset," *IEEE Transactions on Cloud Computing,* 2017.

[31] H. Zhu, *et al.*, "Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM," *IEEE journal of biomedical and health informatics,* vol. 21, pp. 838-850, 2017.

[32] X. Zhang, *et al.*, "Proximity-aware local-recoding anonymization with mapreduce for scalable big data privacy preservation in cloud," *IEEE transactions on computers,* vol. 64, pp. 2293-2307, 2015.

[33] X. Xiao and Y. Tao, "Anatomy: Simple and effective privacy preservation," in *Proceedings of the 32nd international conference on Very large data bases*, 2006, pp. 139-150.

[34] E. Aghasian, *et al.*, "Scoring Users' Privacy Disclosure Across Multiple Online Social Networks," *IEEE Access,* vol. 5, pp. 13118-13130, 2017.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

INFORMATION EXTRACTION FROM DEPENDENCY GRAPHS

A.MURALI KRISHNA
Research scholar, Dept. of Computer Science & Engineering
Rayalaseema University Kurnool – 518007 . (A.P), INDIA.
aitamuralikrishna@gmail.com

Dr. T. SWARNA LATHA
Professor, Dept. of Computer Science & Engineering
Narayana Engineering College Nellore .AP, INDIA.
proftswarnalatha@gmail.com

Abstract— One of the primary goals of biological NLP, and a pre-requisite for practical text mining, is automatic information extraction. This refers to the process of translating a human-readable corpus into structured data for visualization, querying and mining, or indeed any other computational process. Although a rich syntactic representation of a sentence is not necessarily required for IE, this paper presents an experiment to test the hypothesis that using dependency graphs as an intermediate stage can facilitate the extraction of biological interaction from parse trees and provide a powerful and flexible pipeline from raw text to semantic relations.

The experiment was initially performed in three parts, with three different solutions to the same problem. Each one was developed after the previous had been evaluated and all new results had been analyzed, so I will present them chronologically, after describing the LLL Challenge (Learning Language in Logic) which provided the opportunity.

Keywords: Graphs, LLL, NLP, IE, Biological, Chunker, GENIA.

I. INTRODUCTION

The GENIA corpus has been widely used in biological NLP evaluations because it contains various different kinds of linguistic annotation—parts of speech, entity names and types, and constituent structure. However, it is not immediately useful in testing IE algorithms as there is no annotation of the relationships between the entities in each sentence, and this is what is usually required of IE applications in this domain. In order to provide a competitive benchmark for such systems on transcription in *Bacillus subtitles*. Each sentence was annotated by the organizers with a set of ordered tuples representing the pairs of interacting genes and proteins described there in; the causal agent and target of each interaction was marked as such. A test set was also provided, with sentences on the same subject, but with the annotations with held.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The goal of the challenge was to retrieve the interactions described in the test set, using an algorithm trained on the training set. For example, given the sentence *Both SigK and GerE were essential for ykvP expression*, the correct answer would con-sist of two tuples, where the first entity in each tuple is the agent: *SigK ykvP* and *GerE ykvP* . The interactions in the training set fell into three categories: genetic regulatory relationships where no physical mechanism is specified (68%), direct physical interactions such as promoter binding (25%), and relationships implied by member-ship of a region (7%). According to the task guidelines, these relative proportions were similar in the test set, but 50% of the test set sentences had no interactions in (unlike the training set where every sentence had at least one).

II. METHODS

The algorithm was designed to make an initial pass by graph traversal, and then three recovery passes to deal with cases where the initial pass found a likely interaction but failed to assign entities. The aim was to achieve high-precision results with the first pass, and then increase recall with each successive pass until the desired balance between recall and precision was reached. As many design decisions as possible were parameterized rather than hard-coded, and the LLL training data was used instead as a tune set to pick an optimal (or near-optimal) set of parameters to use on the test set. The tradeoffs between precision and recall are a key concept in information extraction, and many of the algorithm's parameters were designed with this balance in mind. For example, a parameter like **remove Negatives** (see below) would be expected to restrict the number of candidate entity pairs that are marked as interacting and thus tend to increase precision at the risk of reducing recall.

Finally, some constraints were introduced in response to common causes of error that were identified during the early development of the algorithm, for example **banSymmetricalSubgraphs** and **remove Negatives**. These were parameterized rather than hard-coded as it is difficult to predict how they will interact with the other parameters and whether they will work best if applied repeatedly or simply checked at one or two key points in the process. As we will see, the amount of linguistic input into the design of this algorithm was fairly minimal, with the structure of the graphs much more important than the labels on the arcs, and the POS tags ignored completely. Rather, the idea was to get a prototype system working to test the basic principles of the dependency graph framework and to use as a frame of reference for other techniques.

Algorithm I's approach makes extensive use of concept of *interaction sub graphs* (see Figure 1 for examples):

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

> "The interaction sub graph for an interaction between two proteins and in a dependency parse is the minimal connected sub graph of that contains, and the word or phrase that states their interaction."

In this task, most of the interacting entities are genes rather than proteins, but the principle is the same, although I did further constrain the definition to include only those where the interaction word is at the root. I constructed a lexicon of interaction words that are likely to indicate that an interaction is being described, based on manual inspection of the training set and extrapolation based on past experience. This included verbs (in various forms) such as bind, phosphorylates and stimulated nouns such as expression, target and repressor, and various other keywords such as dependent.

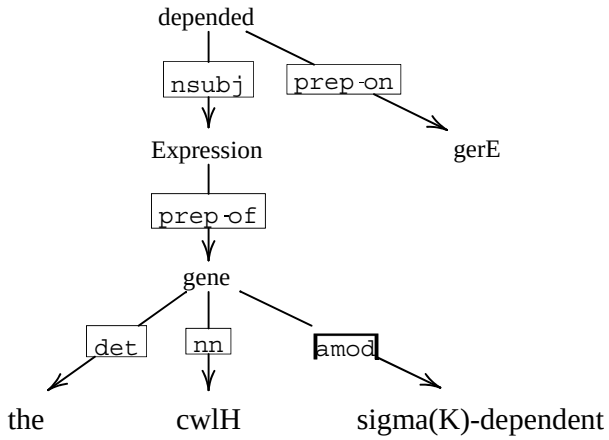


Figure 1: This graph of the sentence Expression of the sigma(K)-dependent cwlH gene depended on gerE contains several candidate interaction subgraphs, which overlap .

2.1 **ALGORITHM I ALSO MAKE USE OF THE FOLLOWING CONCEPTS:**

Interaction triplet: The interaction triplet for an interaction between two genes/proteins (agent) and (target) via an interaction word is the ordered tuple.

Dependency path: A route from node to node in a dependency graph of a sentence, consisting of all the arcs and nodes between and (inclusive) in order.

Note that the arcs in a dependency graph are directional . The direction from parent node to child node is referred to hereafter as 'downstream'.

Contradiction: 'contradiction' refers specifically to cases where the same pair of entities has been marked as interacting twice, with one entity as the agent in one case, and the other as the agent in the other case.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

First pass

The first pass is the core of the algorithm, and is required for a minimum base level of operation. It begins by locating interaction words in the sentence. Then for each of these, the algorithm traverses the dependency graph in the downstream direction in order to find any entities that can form candidate interaction sub graphs with the keyword (as described in Figure 1). If suitable entities are found, this is taken as evidence that a genuine interaction is being described in the sentence, and the entities become potential agents and targets for the interaction.

The details of this process are governed by the following parameters.

deepSearchAgentsFirstPass: If set, an entity whose path from the interaction word contains a dependency type characteristic of the agent role *at any point* is classified as an agent.

lateSearchAgentsFirstPass: If **deepSearchAgentsFirstPass** is not set, and this Parameter is, the algorithm requires that a dependency type characteristic of the agent role is present on the last arc in the path from the interaction keyword to a given entity, in order for that entity to qualify as an agent**.**

If neither of the above parameters are set, the algorithm only considers the dependency type of the first arc in the path from the interaction keyword to the entity, when choosing agents for the interaction keyword. This is illustrated in Figure 4.2.

deepSearchTargetsFirstPass: If set, an entity whose path from the interaction word contains no dependency types characteristic of the agent role *at any point* is classified as a target.

lateSearchTargetsFirstPass: If **deepSearchTargetsFirstPass** is not set, and this parameter is, the algorithm requires that no dependency type characteristic of the agent role is present on the last arc in the path from the interaction keyword to a given entity, in order for that entity to qualify as a target.

As with **deepSearchAgentsFirstPass** and **lateSearchAgentsFirstPass**, if neither of these parameters is set, the critical arc on the path to a candidate target is the nearest one to the interaction keyword. However:

banAgentsAsTargets: If this is not set, the state of the previous two parameters is ignored, and any entity downstream of the interaction keyword is a valid target, regardless of the presence of agent-like dependencies in the paths from the keyword.

It should be clear then that picking targets for an interaction is easier than picking agents, as the dependency types that characterize the agent role are in a minority, and the target selection process has the option of performing no filtering at all on dependency types. The design of the second pass (see below) reflects this fact.

haltSearchAgentsFirstPass: If set, only the first agent entity found along a given dependency path will be admissible. If not set, multiple agents along the same path will be allowed.

haltSearchTargetsFirstPass: The counterpart of the previous parameter; if set,

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

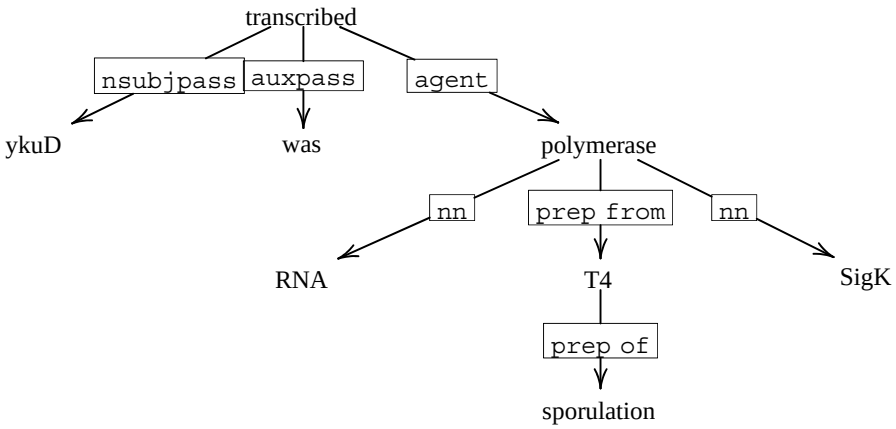


Figure 2: This graph of the sentence ykuD was transcribed by SigK RNA polymerase from T4 of sporulation illustrates the agent selection process in the first pass.

However, if **lateSearchAgentsFirstPass** is set (and **deepSearchAgentsFirstPass** isn't) the algorithm will only look at the dependency *immediately governing* each candidate entity when trying to assign the agent role. In this case, that is nn (noun compound modifier), which is not characteristic of the agent role; the first pass will thus be unable to determine which of the entities is the agent. Note that there is an error in this graph, because the parser has chosen to attach fromT4ofsporulation to polymerase, and not to transcribed, which it really modifies; fortunately this does not affect the algorithm's interpretation.

only the first target entity found along a given dependency path will be admissible. If not set, multiple targets along the same path will be allowed. See Figure 4.3 for an illustrative example.

filterTripletsAfterFirstPass: If set, the global filtering criteria defined by **banIdentityTriplets**, **banSymmetricalSubgraphs** and **removeNegatives** will be applied after the first pass completes.

Second pass

The second pass operates on those interaction keywords which have multiple candidate targets from the first pass, but for which the first pass has been unable to identify an agent. It was introduced because the criteria in the first pass for selecting agents are much narrower than those for identifying targets, and parse errors and complex

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

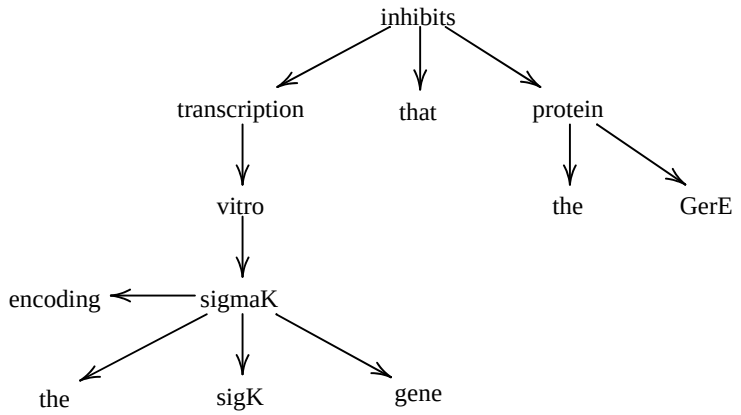


Figure 3: This graph of the sentence the GerE protein in hibits transcription in vitro of the sigK gene encoding sigmaK demonstrates the principles of the **haltSearchTargetsFirstPass** parameter. The dependency types have been left off this diagram for simplicity because they are not relevant here; assume that the algorithm has already identified GerE as the agent of inhibits, and sigK and sigmaK as targets. However, if **haltSearchTargetsFirstPass** is set, then sigK will be disallowed from acting in this role, as to get to it from inhibits by graph traversal is impossible without crossing sigmaK which would be a breach of the halting condition. In this case, this strategy makes little practical difference; since the LLL dataset does not distinguish between genes and their products, sigK and sigmaK are interchangeable, but such things are not considered at the graph traversal stage. There are two significant but not catastrophic parse errors evident in the structure of this graph, but as they do not make this demonstration of search halting any less valid, identifying them is left as an exercise to the reader.

phrase structures can exacerbate this. Based on the following parameters, the algorithm attempts to reassign one of the candidate targets as an agent, thus creating a new interaction triplet between the new agent and each of the remaining targets. **doSecondPass**: This controls whether the second pass is run at all.

agentSelectorSecondPass: Allowable values are the same as for **agentSelector-ContradictionResolver**,togetherwith values GRAPHFURTHESTand GRAPHNEAREST which refer to the distance in arc ‘hops’ from the interaction word. This parameter is used to select an agent from the candidate targets (ties are broken randomly), provided the criterion specified by the next parameter is met. **agentRestrictorSecondPass**: Allowable valuesare BEFORE,AFTER andEITHER. If this is set to BEFORE or AFTER, the candidate agent will be admissible only if it is before or after the interaction word in sentence word order respectively. If the entity selected according to the previous parameter is inaccessible because of this one, no agent is returned; the selected entity is returned to the pool of targets. **filterTripletsAfterSecondPass**: If set, the global filtering criteria defined by

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

banIdentityTriplets, **banSymmetricalSubgraphs** and **removeNegatives** will be applied after the second pass completes.

Third pass

The third pass operates on those interaction words which have at least one candidate target from the first pass, but for which the first and second passes have been unable to identify an agent. The algorithm attempts to identify a plausible agent from *any* of the entities in the sentence, apart from those which are already marked as targets for this interaction word, regardless of whether or not they are accessible by downstream graph traversal from the interaction word. The process is governed by the following parameters.

doThirdPass: This controls whether the third pass is run at all.

agentSelectorThirdPass: Allowable values are WORDFURTHEST, WORDNEAREST, LEFTMOST and RIGHTMOST. GRAPHFURTHEST and GRAPHNEAREST are not allowed because graph traversal is not used in this pass.

agentRestrictorThirdPass: Allowable values are BEFORE, AFTER and EITHER, as in **agentRestrictorSecondPass**. As before, if the entity selected according to the previous parameter is overruled by this parameter, no agent is returned.

filterTripletsAfterThirdPass: If set, the global filtering criteria defined by **banIdentityTriplets**, **banSymmetricalSubgraphs** and **removeNegatives** will be applied after the third pass completes.

Fourth pass

The fourth pass is a final fallback stage for interaction words for which neither an agent nor a target has been identified by any of the previous passes. It attempts to assign an agent and a target from anywhere else in the sentence, regardless of graph connectivity, based on the following parameters.

doFourthPass: This controls whether the fourth pass is run at all.

agentSelectorFourthPass: Allowable values are WORDFURTHEST,WORDNEAREST, LEFTMOST and RIGHTMOST, as in **agentSelectorThirdPass**.

agentRestrictorFourthPass: Allowable values are BEFORE,AFTER and EITHER as in **agentRestrictorThirdPass**.

targetSelectorFourthPass: This parameter takes the same values as **agentSelectorFourthPass**, but governs the search for a target of the interaction rather than an agent.

targetRestrictorFourthPass: This parameter takes the same values as **agentRestrictorFourthPass**, but again governs the search for a target. As in the previous two passes, if a decision made in accordance with a selector parameter is overruled by its corresponding restrictor parameter, no entity is chosen for the role in question.

filterTripletsAfterFourthPass: If set, the global filtering criteria defined by

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

banIdentityTriplets, **banSymmetricalSubgraphs** and **removeNegatives** will be applied after the fourth pass completes.

III. Overview of experimental protocol

A high-level overview of the structure of this experiment is presented in Figure 4.4 for the parameter optimization process using labeled (training) data, and Figure 4.5 for the testing process on unseen (test) data. The latter process is much simpler, being essentially a linear series of steps that incorporates the results (optimal parameter sets) from the former process.

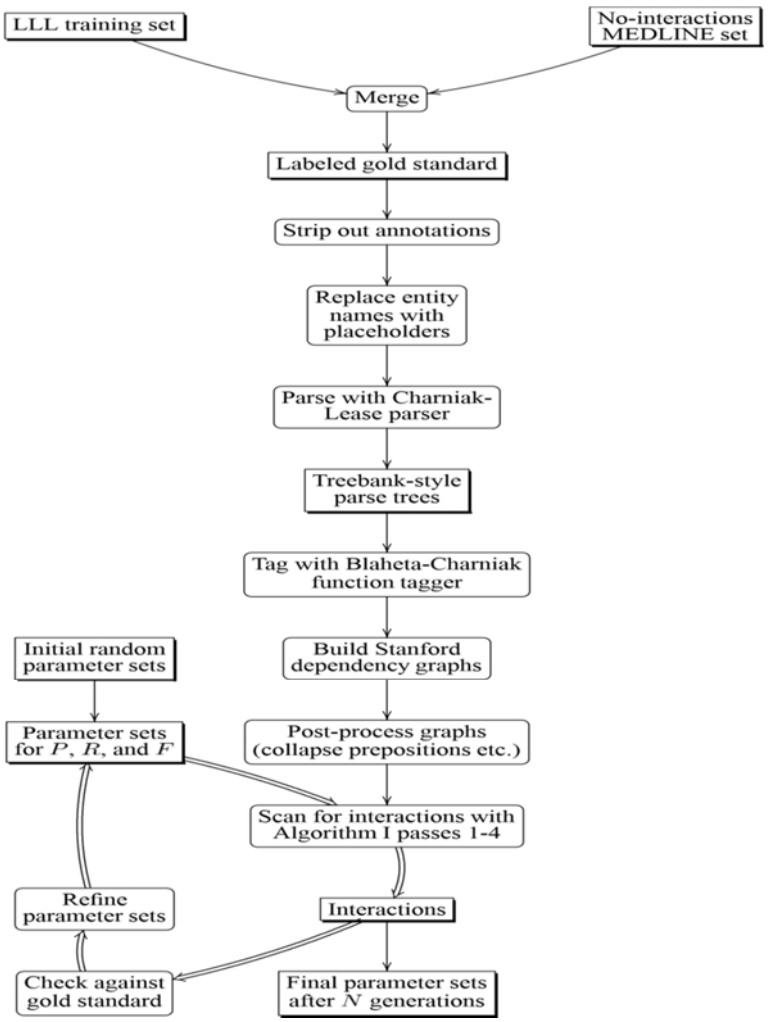


Figure 4: Overview of data preparation and parameter optimization protocol for Algorithm I. Shaded rectangles represent data and rounded boxes are actions. The lines in the training cycle are doubled to indicate iteration; the cycle repeats once every generation under the control of the genetic algorithm.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

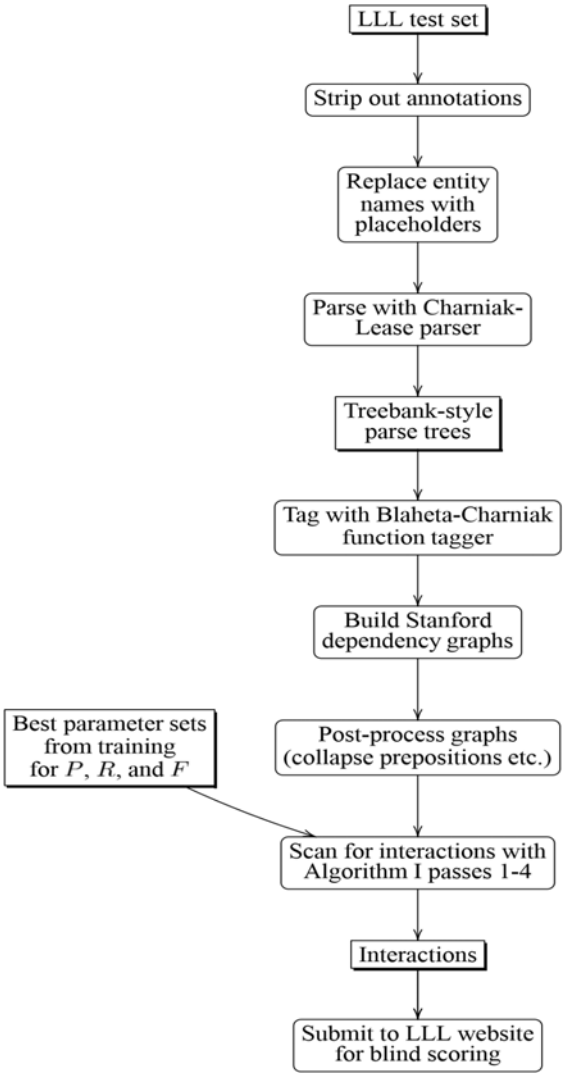


Figure 5: Overview of data preparation and evaluation protocol for Algorithm I. Shaded rectangles represent data and rounded boxes are actions.

Scores on LLL corpus for Algorithm I						
Parameter set	Training set			Test set		
high-P	**71.9**	25.8	38.0	**85.7**	21.6	34.6
high-R	32.3	**77.4**	45.6	31.4	**61.4**	41.6
high-F	58.9	60.4	**59.6**	57.3	46.9	**51.6**

Table 1: Algorithm I's relationship extraction scores (precision, recall and F-measure)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Each of the three top parameter sets chosen by JGAP—hereafter referred to as *highP*, *high-R* and *high-F*— were then used to process the LLL test set, and the results were submitted to the LLL scoring service for evaluation. Table 1 shows the scores achieved on the LLL training set (during parameter selection) and test set, on the hardest subtask. Each parameter set has a considerably lower recall on the test set than on the training set; the fall-off was a little larger than I expected, given that the training and test sentences were drawn from the same topic and time period, and even selected so that they have the same distribution of interaction types. The teams in the challenge were not required to report their scores on the training set, which is a shame as that would have been informative. Group 6 mention that their official run which scored= 52.6 was the original highest-scorer on the training set, with = 65; although they don't give the source of the errors, looking at the individual test set scores (= 60.9,= 46.2) suggests that it was also primarily a recall problem. It is likely that there are syntactic constructions in the test set that are not found in the training set—such is the danger of using small datasets. Furthermore, given the big difference between Group 6's results with the perfect Link annotations and the noisy CCG parse, I suspect that some of these are of a kind that is problematic for parsers.

Precision, on the other hand, was much easier to maintain on the test set, and even went up considerably under the *high-P* parameter set to 85.7. There are uses for high precision methods; although a recall of 21.6 means that only one in five interactions are successfully recovered, this is less of a problem in large datasets where the better established claims will be repeated several times any way .

Table 2 shows the results for each of these parameter sets—hereafter referred to as *high-P*, *high-R* and *high-F*—in context with the original LLL contestants, ranked once again by F-measure. *high-P* is near the bottom, but *high-R* and *high-F* both come in above the only other run to handle both sentence types without using the linguistic annotations. Although they are still quite far from Group 6's high scorers, those rely heavily on the perfect Link data. Comparing like for like, then, *high-R* and *high-F* were very successful.

Two small points must be made about the rules of the task which may have had a minor negative impact on these scores. One is that in cases where a sentence reports two distinct interactions between the same two entities, each one is annotated separately, so there are potentially two true positives available. For example:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Results of the LLL Challenge, plus Algorithm I					
Sentences	Annotations	Group			
easy	yes	6a	65.0	72.2	**68.4**
both	yes	6b	63.2	66.2	64.7
both	yes	6c	55.6	53.0	54.3
easy	no	6d	68.5	44.4	53.9
easy	yes	6e	60.9	46.2	52.6
easy	no	1	50.0	53.8	51.8
both	no		57.3	46.9	51.6
easy	yes	4	37.9	55.5	45.1
both	no		31.4	61.4	41.6
both	no	6f	50.0	33.7	40.2
easy	no	5a	25.0	81.4	38.2
both	no		**85.7**	21.6	34.6
easy	yes	5b	20.5	90.7	33.4
both	yes	3	51.8	16.8	25.4
hard	yes	5c	14.0	93.1	24.4
hard	no	5d	14.0	82.7	24.0
easy	no	2	10.6	**98.1**	19.1

Table 2: The results of the LLL Challenge, in descending order of F-measure. indicates unofficial runs. , and are my *high-P*, *high-R* and *high-F* runs.

IV. CONCLUSION

However, the strength of dependency-based methods in general was reinforced by RelEx's striking results. It is hard to do a detailed comparison without access to RelEx itself (it is not publically available), and some of the advantage may be due to an ab-124sence of hard sentences in its test data, but there are interesting aspects to its design which seemed like they might be important. Firstly, noun phrase chunking seems like a sensible way to reduce unnecessary variability in the graphs and thus make simple methods more effective. In a phrase like X inhibited Y expression the target of the interaction is not attached directly to the verb as an object—expression is the object node in the graph, and Y is attached to it as a pronominal modifier. However, if a noun phrase chunker is applied so the phrase becomes X inhibited Y expression then a simple pattern or routine designed to capture X inhibited Y will work here too. This means that the system must detect entities that are substrings of nodes, but this is already necessary as the Charniak-Lease parser treats strings like sigma(H)-dependent as single words so they end up in one graph node.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

REFERENCE

[1]. W. W. Chapman and K. B. Cohen, “Current issues in biomedical text mining and natural language processing,” Journal of Biomedical Informatics, vol. 42, no. 5, pp. 757–759, 2009. View at Publisher · View at Google Scholar · View at Scopus.

[2]. M. E. Cusick, H. Yu, A. Smolyar et al., “Literature-curated protein interaction datasets perspective,” Nature Methods, vol. 6, pp. 39–46, 2009. View at Publisher · View at Google Scholar

[3]. R. A.-A. Erhardt, R. Schneider, and C. Blaschke, “Status of text-mining techniques applied to biomedical text,” Drug Discovery Today, vol. 11, no. 7-8, pp. 315–325, 2006. View at Publisher · View at Google Scholar · View at Scopus

[4]. D. Zhou and Y. He, “Extracting interactions between proteins from the literature,” Journal of Biomedical Informatics, vol. 41, no. 2, pp. 393–407, 2008. View at Publisher · View at Google Scholar · View at Scopus.

[5]. A. Airola, S. Pyysalo, J. Björne, T. Pahikkala, F. Ginter, and T. Salakoski, “All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning,” BMC Bioinformatics, vol. 9, no. 11, article 52, 2008. View at Publisher · View at Google Scholar · View at Scopus

[6]. H. Kilicoglu and S. Bergler, “Adapting a general semantic interpretation approach to biological event extraction,” in Proceedings of the BioNLP Shared Task Workshop, pp. 173–182, 2011.

[7]. B Hari Krishna, S Kiran, G Murali, R Pradeep Kumar Reddy “Security issues in service model of cloud computing environment” Procedia Computer Science, Elsevier 246-251, 87 2016.

[8]. W. A. Baumgartner Jr., K. B. Cohen, and L. Hunter, “An open-source framework for large-scale, flexible evaluation of biomedical text mining systems,” Journal of Biomedical Discovery and Collaboration, vol. 3, article 1, 2008. View at Publisher · View at Google Scholar · View at Scopus.

[9]. B. Harikrishna, S. Kiran, R. Pradeep Kumar Reddy, Protection on sensitive information in cloud — Cryptography algorithms IEEE digital Library 10.1109/CESYS.2016.7889894.

[10]. Y. Garten, A. Coulet, and R. B. Altman, “Recent progress in automatically extracting information from the pharmacogenomic literature,” Pharmacogenomics, vol. 11, no. 10, pp. 1467–1489, 2010. View at Publisher · View at Google Scholar · View at Scopus.

[11]. H. Liu, R. Komandur, and K. Verspoor, “From graphs to events: a subgraph matching approach for information extraction from biomedical text,” in Proceedings of the BioNLP Shared Task 2011 Workshop, pp. 164–172, 2011.

[12]. O. Frunza and D. Inkpen, “Extraction of disease-treatment semantic relations from biomedical sentences,” in Proceedings of the Workshop on Biomedical Natural Language Processing (ACL '10), pp. 91–98, Uppsala, Sweden, July 2010.

[13]. A. Sharma, R. Swaminathan, and H. Yang, “A verb-centric approach for relationship extraction in biomedical text,” in Proceedings of the 4th IEEE International Conference on Semantic Computing (ICSC '10), pp. 377–385, IEEE, Pittsburgh, Pa, USA, September 2010. View at Publisher · View at Google Scholar · View at Scopus.

[14]. B. Harikrishna, N.Anusha, K.Manideep, Madhusudhanarao, Ch “Quarantine Stabilizing Multi-Keyword Rated Discover with Unfamiliar ID Transferover Encrypted Cloud Warning” IJERCSE Vol 2, Issue 2, February 2015.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Object Detection in Fog Degraded Images

Gurveer Singh and Dr. Ashima Singh

Thapar Institute of Engineering and Technology, Patiala, India

{gurveer7393@gmail.com, ashima@thapar.edu}

Abstract – Image processing is the technique to fetch the valuable data from the given images for different purposes like improvement in visualization of an image and to measure structure or features from the extracted data. High-quality images and videos are easy to predict, and classify, whereas detecting hazy or foggy image is a cumbersome issue. This paper has proposed an efficient methodology integrating various techniques of image processing like Discrete Wavelet Transform (DWT) and Convolutional Neural Network (CNN) for defogging images with prior pre-processing using guided filter. The proposed technique has improved the related standard performance metrics like PSNR, MSE and IIE significantly.

Keywords – DWT (Discrete Wavelet Transform), Guided Filter, Image processing, Enhancement Techniques, visibility and contrast.

I. Introduction

Images captured in different climates or atmosphere exhibits the poor visibility results because of haziness, fog, or blurriness in an image and requires extra efforts for the meaningful results. Image processing is vital method that allows real time and post processing techniques on images. Outdoor captured images are significantly devastating due to the suspended particles like dust, water droplets in air, low light etc. [1].

(a) (b)

Figure 1. Image quality: (a) foggy image, (b) poor quality image [2]

Such situation makes detection of objects difficult and targets unclear in video processing such as extraction of features, tracking of the target, and recognition of obstacles, which may lead to the numerous accidents on the roads, oceans, or in the air as the result of low visibility. Hence, there is a need to design a method that enhances the visibility of an image.
An image is a matrix or an array, with number of pixels, which are elements of an image and are arranged in the form of rows and columns. Digital and analogue are the two types stated

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

by image processing. The foremost merits of digital image processing are repeatability, adaptability, and protection of actual data precision. Numerous methods of processing an image are image pre-processing, image filtration, image segmentation, image classification, and recognition.

II. Methodology

The proposed work includes following steps: i. image Acquisition; ii. Grayscale Conversion; iii. Brightness Mapping; iv. Transmission and Estimation Process; v. Guided Filter; vi. Wavelet Transformation (DWT); vii. Convolutional Neural Network(CNN); viii. Performance Parameters; ix. Comparison

As shown in Figure 2 and Figure 3, the proposed work is sectioned in two phases: Phase I and Phase II. In Phase I, we have selected an image from the dataset , (dataset from the UCI Machine Learning Repository).

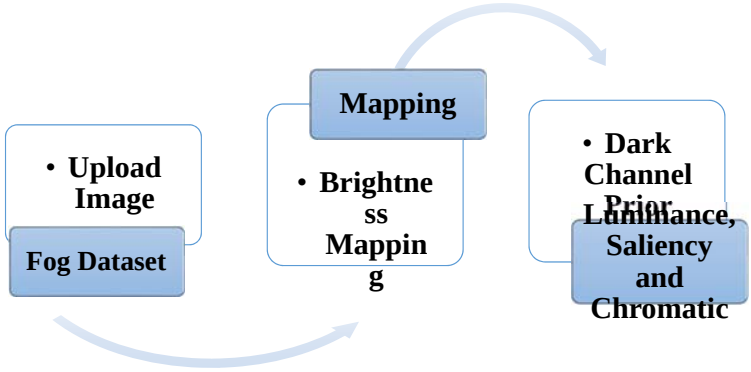


Figure 2: Phase I

Figure 2 describes the initial Phase I in which fog deformed image data set is used to upload an image and transformation is performed using brightness mapping. Dark channel prior is obtained in the form of Luminance, Saliency and Chromatic as an output.

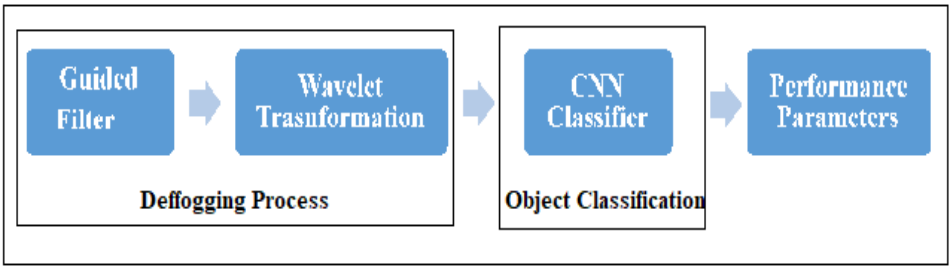


Figure 3: Phase II

Figure 3 illustrates the final Phase II in which defogging of an image through guided and wavelet transform is implemented and CNN classifier classifies an object.

III. Related Work

N. Sangeetha, et al., [2] observed different techniques of enhancement methods to defog the image. It compared the results based on PSNR (Peak Signal to Noise Ratio), NPCR (The number of changing pixel rate), IQM (Image correlation coefficient), and BAI (Blind Assessment Indicator). Data set is

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

used related to forest data and road scene only. Results showed that SWT method is the providing the best results as compared to the DCT (Discrete Cosine Transform), Weiner Filtering, Holomorphic and high boost filtering. DCT had high NPCR only for first image whereas IQM and holomorphic had high value for the both images i.e. original and filtered but contrast rate was high in second image for holomorphic.

Kaiming He, [3] implemented a work in the image prior i.e. dark channel prior using an image as input to remove the haziness from image. Fog free images have pixels with low intensity at least a single colour channel. Thickness of the fog is identified using prior method and results in fog free image with high resolutions. Author also compared the results of his images with related work and this single dark channel prior was the best resulted.

Baojun Qi et al., [4] developed a rapid method to improve the visibility performance of foggy views with a single image. It has considered the thickness of the fog image with Gaussian filter and an intensity image which is composed of the noisy image's DCCs (Darkest colour Channel). It is depends on the observation the regions, edges of the air light were same to that of the noisy image minimum frequency component. Now next, to manage mis-estimates of the air light a revised picture model and a scenario of the nested multiple image windows are implemented to improve the low visibility and make faded objects in the distance to be show.

Kristofor B. et al., [5] presented single image defogging approach that uses a new method to calming the estimate of amount of noise in an image with the LAWF (Local Adaptive Wiener Filter). It provided a answer of estimating fog performance parameters for the filter when the study and fog are co-related by de-correlating with the estimated de-fogged image.

Wei Song et al., [6] presented a real time video de-fogging approach based on the context sensitiveness by using enhanced GFA (Guided Filtering Algorithm) and enhancing the single image frame de-fogging effect within a constant computation interval of time period. The main advantage of contextal data, proposed various strategy integration video de-fogging approach experimental consequences define that the approach is able to de-fog in real-time moving camera videos in calculate the de-fogging performance metrics.

IV. PROPOSED MODEL

In this section, for enhancing the performance and for accurate results, some techniques and algorithms are used namely: - Discrete Wavelet Transformation and guided filter.

A. Discrete Wavelet Transformation: In digital image processing the compression is the best technique and successful in field of digital images. There are various methods are used for compression. But the better option is Discrete Wavelet Transforms, also successful in signal processing. More than that it is high efficient and flexible for decompose signals [8]. Basic Functions in Discrete Wavelet Transformation: In DWT are two basic functions i.e. HAAR and DAUBECHIES wavelets. Haar Transformation: It discovered by Hungarian Mathematician Alfred Haar. It is discontinuous and a step functions. Particularly, it is preferred for ortho-normal systems for square integer able function. It is simple transformation. Haar uses averaging and differencing terms [9]. Some significant to note the following properties:

i. Wavelet methods are spatially localized.
ii. Wavelet methods are dilated, translated and scaled versions of normal wavelet and
iii. Every set of wavelet methods forms an orthogonal set of normal methods. [10]
iv. Images converted into wavelets that are more efficient than pixel blocks.
v. In DWT no input coding is required for overlapping.

vi. It is good for localization time and frequency.

B. Guided Filter Method: Guided Filtered Image de-fogging approach is used in this section for individual frame image and this approach is depends on DCP (Dark Channel Prior) information approach [11]. The computer vision areas are following the foggy image de-gradation structure eq in world-wide used: -

$$I(x1) = j(x1).t(x1) + A(1-t(x1)) \text{ ..... (i)}$$

Where I (x1) is the recent image to be de-fogged, j(x1) is fog free images to be improved, A is global atmospheric light component, t(x) is the transmission rate. Now I(x1) is called situation and j(x1) is the compulsory target value, it is clear i.e., is an eq ii with solutions, therefore, we required priori information.

Formula (i) transforming, we get:

$$\frac{Ic(x1)}{Ac} = t(x1)\frac{j(c)(x1)}{Ac} + 1 - t(x) \text{ ..... (ii)}$$

$$\min y \in \Omega(x1)\left(\min c \frac{Ic(x1)}{Ac}\right) = t(x) \min y \in \Omega(x1)\left(\min c \frac{jc(y)}{Ac}\right) + 1 - t(x1) \text{ .... (iii)}$$

Where, j dark (x1) = min y $\in$ Ω (x1) (min c $\in$ {r, g, b} jc (y)) is DC (dark Channel), Jc it means individual channel for a color picture and Ω (x1) means a window centered on x1. Theory of the DCP , in most non-local edge of the sky, where there is always at least individual channel that has a minimum value for some of image pixels. In other words, the decreasing value of the light intensity of the area is minimum number. The decreasing number of the intensity of light of area with fog is a higher value. The DCP theory defines j dark --> 0.

DCP theory into application and transforming the formula (3) by taking into account the required to retain a certain degree of fog, then we get:

$$t(x1) = 1 - \omega \min y \epsilon \Omega(x1)\left(\min c \frac{Ic(y)}{Ac}\right) \text{ ..... (iv)}$$

Where, ω means the degree of fog, which is given 0.95 by experience.

C. Deep learning

Deep Learning is a famous and intensively used approach in machine learning. CNN refers to conventional neural networks that are a sub category of DNN (deep neural networks). CNN can access both forward and backward propagation. Hubel and Wiesel do the discovery of conventional neural network in early 1960's. It is discovered while searching on neurons mainly approached for sensitive orientation of visual systems of cats. The implementation of CNN based on two primary layers as FE (feature extraction) and other one is FM (feature maps). Basic Layers in CNN are:

i. Feature Extraction Layer: In this layer, the neurons input is linked to local fields and captures all required features and then links to other neurons are created automatically.

ii. **Feature Maps:** Every layer contains a feature map, which is a plane. The weights of every neuron are similar in this layer. Its framework is dependent upon the sigmoid function.

D. Workflow of the proposed technique

The proposed work is explained in given steps: -

STEP I: Image Acquisition: This is the initial Phase I . In this process, the data set is collected from the UCI Machine Learning repository site. Unclear and noisy images like foggy images, roadside scenes and cloudy images from the dataset folder. Image taken from the repository is shown in Figure 4.

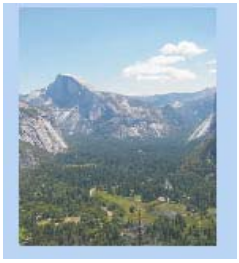
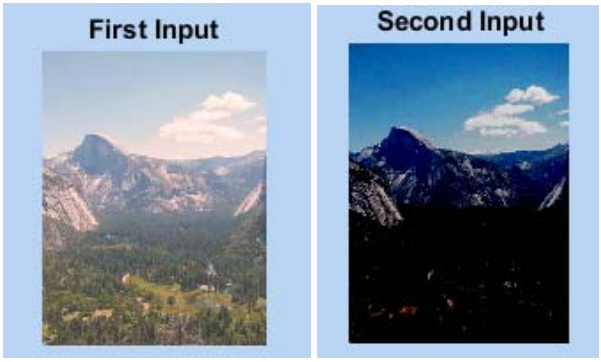


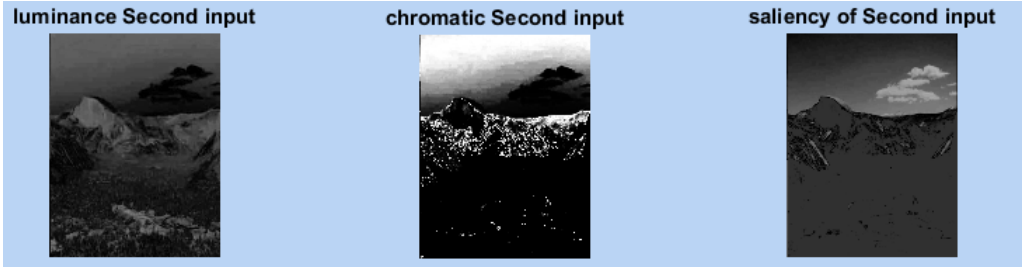
Figure 4: Original Image

STEP II: Pre-processing and Mapping: In this pre-processing phase, convert the original image into a grayscale image as shown in Figure 5. Fog image recorded by sensors on a satellite contain errors in regard to the brightness values of the image pixels. To extract the colour components based on brightness mapping methods, extract the views in transmission and estimation fogging image. Implemented the guide filter in the de-fogging image and clearly found the features i.e. luminance, saliency and chromatic feature detect.

(a)



(b)



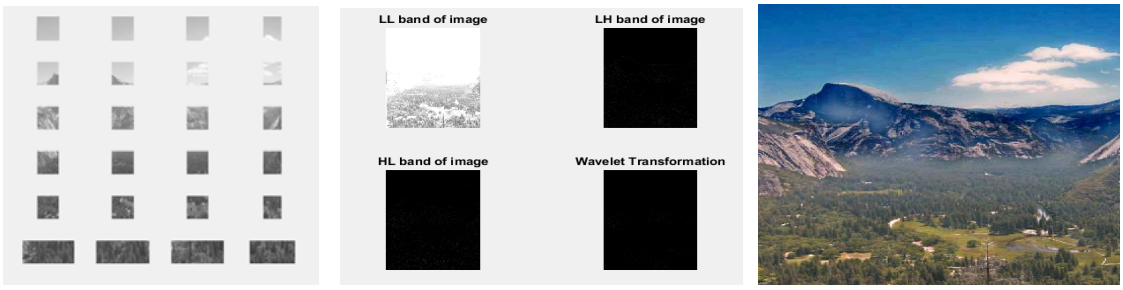
(c)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

(d)

Figure 5: (a) Grayscale Image (b) Brightness Mapping Image (c) Transmission and Estimation Image and (d) Enhanced Image.

STEP III: Filtration: Image shows that the fogging image and another enhanced colour image. The novel implemented method divides the image into four frames i.e. LL, HL, HH, LH bounds and noise removal using discrete wavelet transformation filtered image produces, see Figure 6. These filter methods are used for removal of interference in the fogy image. DWT algorithm using HAAR WAVELET is discontinuous. Particularly, it is preferred for ortho-normal systems and for square integer able function. HAAR uses averaging terms to eliminate data. The discrete signal in HAAR decomposes into two sub-signals of half of its length.



(a) (b) (c)

Figure 6: (a) Block wise division Image, (b) Band Conversion and (c) Inverse Discrete Wavelet Transformation Image

STEP IV: Classification and Detection: It refers to a sub-section of ML (Machine Learning) that depends upon learning levels of representations, equivalent to a hierarchy of characteristics in which the high-level concept is defined from low level and the similar concept could help to explain various high-level. It helps to understand the information like as a text, audio and images. Deep Learning comes from the analysis ANN, MLP which includes more HL (Hidden Layers) is a DLS (Deep Learning Structure). Normally, the design of deep learning algorithm adds binary layers one of them is feature extraction layer. The initial of individual neuron is associated with the local receptive areas of the existing layer and fetch the local characteristic. The local feature extracts the location relation with itself and other characteristics are also determined. Other is the characteristic map layer which is an individual computing layer of the system and is composed of a plurality of characteristic map. All of the feature-map is a plane and the weight of the neurons in the plane is equivalent. Design of feature-map utilizes the SIGMOID method as an activation method of the CNN, which creates the feature-map have modified invariance. In addition, the NNs in

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

the similar mapping plane share weights; the various numbers of free metrics of the system is optimized. Each CL (Convolutional Layer) in the CNN is preferred by a computing-layer, which is utilized to evaluate the local average. In this unique binary feature, extraction design optimizes the resolution [48]. The main benefit of the CNN with respect to the NN is the special design of Convolutional Neural Network local shared weights.

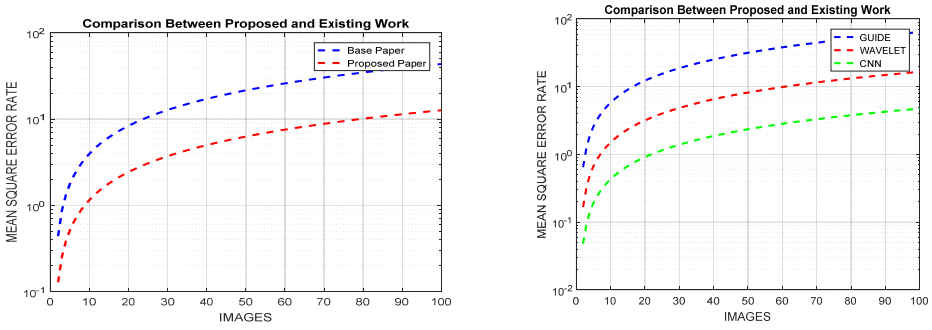
Figure 7: Object Identification

STEP V: Performance Metrics: In this phase the performance parameters like PSNR (Peak Signal to Noise Ratio), MSE (Means Square Error) and IIE (Image Information Entropy) are calculated and compared with the performance of the existing parameters.

IV. Results

The standard performance metrics for the proposed technique is compared with the selected base paper titled Image defogging using enhancement techniques [2]. Standard parameters used for performance evaluation are:

Mean Square Error Rate (MSE): MSE measures the average of the squares of the errors. It is evaluated for the proposed technique; and also has been compared with the existing work[2].



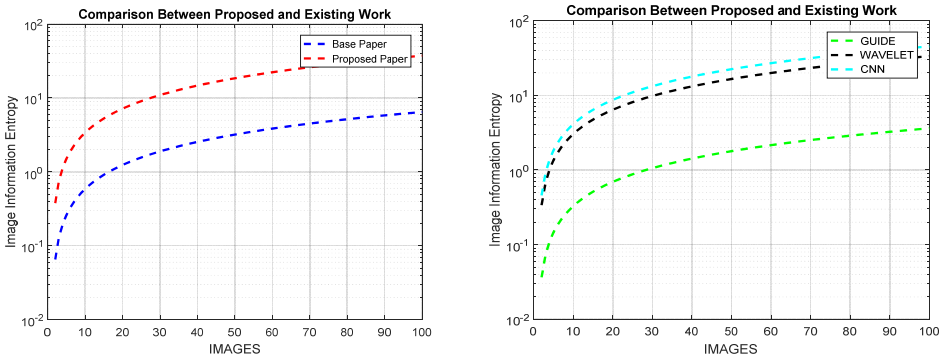
(a) (b)

Figure 8: (a) Comparison (Proposed and Existing work) and (b) Comparison (Techniques).

Image Information Entropy

Figure 9 shows the comparison between proposed and existing works. IIE is a quality which is used to explain the defog image, that is the amount IIE which must be implied for by a deep learning algorithm. The minimum entropy image like those surrounding a lot of Grayscale image, have minimum contrast and large image of pixels with the same values.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

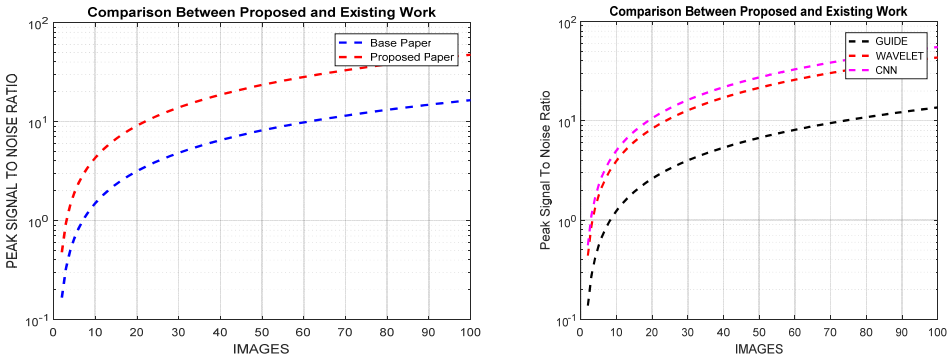


(a) (b)

Figure 9: (a) Comparison (Proposed and Existing work) and (b) Comparison (Techniques).

Peak Signal to Noise Ratio

Figure 10 shows the PSNR ratio, is the ratio between the maximum possible power of an image signal and the power corrupt noise that affects the reliability of its representation.



(a) (b)

Figure 10: (a) Comparison proposed and existing work and (b) Comparison between techniques

Table 1 discusses the performance parameters of three selected methods.

Table 1: Comparison Parameters

Parameters	Guided Filter	Discrete Wavelet Transformation	CNN Method
PSNR (%)	13.61	43.42	55.24
IIE (%)	3.61	33.42	45.24
MSE	64.012	16.5	4.7

Table 2 describe that the comparison between proposed and existing work in various parameters and proposed method.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 2: Performance Parameters comparing existing and proposed method

Parameters	Discrete Wavelet Transformation (Existing work)	CNN and DWT (Proposed Work)
PSNR (%)	43.42	55.24
IIE (%)	33.42	45.24
MSE	16.5	4.7

Table 1 and Table 2 describe that the comparison between proposed and existing work in various parameters and proposed method (DWT and CNN) parameters like as a Means Square Error Rate (MSE), Image Information Entropy(IIE) and Peak Signal to Noise Ratio(PSNR).

V. CONCLUSION

In this paper, several improvement methods to defog the image are discussed. Discrete wavelet transformation method has shown better results for foggy images. Images with huge edges having similarity with atmospheric light for out-door view systems are not easy to detect with existing defogging approaches. In this method, the distant objects are largely improved except for some negligible blocking effects. The approach is expected to be useful in active safety systems. In the proposed method, an enhanced guided filtering approach based on contextual information is implemented which is further integrated with CNN and DWT methods to detect the object. The proposed technique has improved the related performance metrics like PSNR, MSE and IIE significantly.

REFERENCES

1 https://gulfnews.com/news/uae/weather/weather-it-s-a-foggy-4-2-c-in-this-part-of-uae-1.2156411 Available at: https://www.tripadvisor.com/LocationPhotoDirectLinkg268-d1226646-i154158195
2 Sangeetha, N., &Anusudha, K. (2017, January). Image defogging using enhancement techniques. In Computer, Communication and Signal Processing (ICCCSP), 2017 International Conference on (pp. 1-5). IEEE.
3 Kaiming He, Jian Sun, and Xiaoou Tang, “Single Image Haze Removal Using Dark Channel Prior”, IEEE transactions on pattern analysis and machine intelligence, vol. 33, no.12, December 2011.
4 Baojun Qi, Tao Wu, Hangen He. (2009, December). A new defogging method with nested windows. In Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on (pp. 1-4). IEEE.
5 Kristofor B. Gibson, Truong Q. Nguyen. Fast single image fog removal using the adaptive wiener filter. In Image Processing (ICIP), 2013 20th IEEE International Conference on (pp. 714-718). IEEE.
6 Wei Song, Bangfei Deng, Haibing Zhang. An adaptive real-time video defogging method based on context-sensitiveness. In Real-time Computing and Robotics (RCAR), IEEE International Conference on (pp.406-410). IEEE.
7 Kin Gwn Lore, Adedotun Akintayo, Soumik Sarkar**.** LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 61, 650-662.
8 D. Gupta, and S. Choubey. Discrete wavelet transform for image processing. International Journal of Emerging Technology and Advanced Engineering, 4(3), 598-602.
9 A. Cichocki, and S. I. Amari. Adaptive blind signal and image processing: learning algorithms and applications (Vol. 1). John Wiley & Sons.
10 M. J. Shensa. The discrete wavelet transform: wedding the atrous and Mallat algorithms. IEEE Transactions on signal processing, 40(10), 2464-2482.
11 K He, J Sun, X Tang. Single image haze removal using dark channel prior[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2011, 33(12): 2341-2353.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Dr Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Dr. P. Vasant, University Technology Petronas, Malaysia
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Dr. Praveen Ranjan Srivastava, BITS PILANI, India
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Dr. Tirthankar Gayen, IIT Kharagpur, India
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan
Prof. Ning Xu, Wuhan University of Technology, China
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan

CALL FOR PAPERS
International Journal of Computer Science and Information Security

IJCSIS 2018-2019
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500
http://sites.google.com/site/ijcsis/